

Utah State University

DigitalCommons@USU

---

All Graduate Theses and Dissertations

Graduate Studies

---

5-2010

# Multivariate Bayesian Machine Learning Regression for Operation and Management of Multiple Reservoir, Irrigation Canal, and River Systems

Andres M. Ticlavilca  
*Utah State University*

Follow this and additional works at: <https://digitalcommons.usu.edu/etd>



Part of the [Civil and Environmental Engineering Commons](#)

---

## Recommended Citation

Ticlavilca, Andres M., "Multivariate Bayesian Machine Learning Regression for Operation and Management of Multiple Reservoir, Irrigation Canal, and River Systems" (2010). *All Graduate Theses and Dissertations*. 600.

<https://digitalcommons.usu.edu/etd/600>

This Dissertation is brought to you for free and open access by the Graduate Studies at DigitalCommons@USU. It has been accepted for inclusion in All Graduate Theses and Dissertations by an authorized administrator of DigitalCommons@USU. For more information, please contact [digitalcommons@usu.edu](mailto:digitalcommons@usu.edu).



MULTIVARIATE BAYESIAN MACHINE LEARNING REGRESSION FOR  
OPERATION AND MANAGEMENT OF MULTIPLE RESERVOIR,  
IRRIGATION CANAL, AND RIVER SYSTEMS

by

Andres M. Ticlavilca

A dissertation submitted in partial fulfillment  
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Civil and Environmental Engineering

Approved:

---

Dr. Mac McKee  
Major Professor

---

Dr. David Stevens  
Committee Member

---

Dr. Gary Merkley  
Committee Member

---

Dr. Wynn Walker  
Committee Member

---

Dr. Gilberto Urroz  
Committee Member

---

Dr. Byron Burnham  
Dean of Graduate Studies

UTAH STATE UNIVERSITY  
Logan, Utah

2010

## ABSTRACT

Multivariate Bayesian Machine Learning Regression for Operation and Management of  
Multiple Reservoir, Irrigation Canal, and River Systems

by

Andres M. Ticlavilca, Doctor of Philosophy

Utah State University, 2010

Major Professor: Dr. Mac McKee  
Department: Civil and Environmental Engineering

The principal objective of this dissertation is to develop Bayesian machine learning models for multiple reservoir, irrigation canal, and river system operation and management. These types of models are derived from the emerging area of machine learning theory; they are characterized by their ability to capture the underlying physics of the system simply by examination of the measured system inputs and outputs. They can be used to provide probabilistic predictions of system behavior using only historical data. The models were developed in the form of a multivariate relevance vector machine (MVRVM) that is based on a sparse Bayesian learning machine approach for regression. Using this Bayesian approach, a predictive confidence interval is obtained from the model that captures the uncertainty of both the model and the data. The models were applied to the multiple reservoir, canal and river system located in the regulated Lower Sevier River Basin in Utah. The models were developed to perform predictions of multi-time-ahead releases of multiple reservoirs, diversions of multiple canals, and streamflow

and water loss/gain in a river system. This research represents the first attempt to use a multivariate Bayesian learning regression approach to develop simultaneous multi-step-ahead predictions with predictive confidence intervals for multiple outputs in a regulated river basin system. These predictions will be of potential value to reservoir and canal operators in identifying the best decisions for operation and management of irrigation water supply systems.

(131 pages)



To my parents Andres and Maria,  
and my sister Jessica

## ACKNOWLEDGMENTS

First of all, I Thank GOD for his blessing in my personal and professional life.

I would like to express my sincere gratitude to my major professor, Dr. Mac McKee, for giving me the opportunity to work with him. His support, guidance, and encouragement were the keys to the success of this dissertation. I would also like to thank the members of my committee, Drs. Wynn Walker, David Stevens, Gary Merkley, and Gilberto Urroz, for their valuable suggestions.

I really appreciate the help and support of my wife, Inga Maslova. I would also like to thank my officemates Alfonso Torres and Bushra Zaman for useful discussions.

I am grateful to Roger Hansen of the Provo, Utah, office of the U.S. Bureau of Reclamation, and Jim Walker of the Sevier River Water Users Association.

Finally, I would like to express special thanks to my family and friends from Peru for all their help and love.

Andres M. Ticlavilca

# CONTENTS

	Page
ABSTRACT .....	ii
ACKNOWLEDGMENTS .....	v
CONTENTS.....	vi
LIST OF TABLES .....	ix
LIST OF FIGURES .....	x
CHAPTER	
1 INTRODUCTION .....	1
References. ....	6
2 MULTIVARIATE BAYESIAN REGRESSION APPROACH TO FORECAST RELEASES FROM A SYSTEM OF MULTIPLE RESERVOIRS.....	8
Abstract. ....	8
2.1 Introduction.....	9
2.2 Model Description.....	12
2.3 Study Area. ....	15
2.4 Model Application to Multiple Reservoir System. ....	16
2.5 Results and discussion. ....	18
2.5.1 Model Selection. ....	18
2.5.2 Performance evaluation. ....	19
2.5.3 Bootstrap analysis. ....	21
2.5.4 Comparison between MVRVM and ANN. ....	22
2.6 Summary and Conclusions. ....	23
References. ....	25
3 REAL-TIME FORECASTING OF SHORT-TERM IRRIGATION CANAL DEMANDS USING A ROBUST MULTIVARIATE BAYESIAN LEARNING MODEL.....	38
Abstract. ....	38
3.1 Introduction.....	39

3.2	Model Description.....	42
3.3	Study Area. ....	46
3.4	Model Application.....	46
3.4.1	MVRVM forecasting hourly model. ....	47
3.4.2	MVRVM forecasting daily model.....	48
3.4.3	Model Selection. ....	50
3.4.4	Bootstrap analysis. ....	50
3.4.5	Comparison between MVRVM and ANN. ....	51
3.5	Results and Discussion. ....	52
3.6	Summary and Conclusions. ....	54
	References. ....	56
4	REAL-TIME FORECASTING OF STREAMFLOW AND WATER LOSS/GAIN IN A RIVER SYSTEM USING A MULTIVARIATE BAYESIAN REGRESSION MODEL. ....	73
	Abstract. ....	73
4.1	Introduction.....	74
4.2	Material and Methods.....	78
4.2.1	Model description.....	78
4.2.2	Study area and data.....	81
4.2.3	Identification of inputs and outputs.....	83
4.2.4	Model selection. ....	86
4.2.5	Bootstrap analysis. ....	86
4.2.6	Comparison between MVRVM and ANN. ....	88
4.3	Results and Discussion. ....	89
4.4	Summary and Conclusions. ....	90
	References. ....	92
5	SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS. ....	104
5.1	Summary and Conclusions. ....	104
5.2	Recommendations and Future Work.....	108
	References. ....	109
	APPENDICES.....	110
	Appendix A. Kernel Functions.....	111
	Appendix B. Statistics for judging model performance.....	112
	Appendix C. Training phase plots of reservoir release predictions. ....	113

CURRICULUM VITAE.....	116
-----------------------	-----

## LIST OF TABLES

Table	Page
2.1 Selected MVRVM for each type of kernel function.....	27
2.2 MVRVM performance using different statistics .....	28
2.3 Selected ANN model for each type of training function.....	29
2.4 ANN performance using different Statistics. ....	30
3.1 Hourly model performance using different statistics for the testing phase.....	59
3.2 Daily model performance using different statistics for the testing phase. ....	60
4.1 Model performance using different statistics. ....	96
5.1 Overall performance of the MVRVM models.....	106

## LIST OF FIGURES

Figure	Page
2.1 Lower Sevier River Basin and the spatial distribution of the sensor station locations (reservoir releases, streamflows, inflows and outflows). ....	31
2.2 Plot of observed versus predicted releases of Sevier Bridge reservoir one day ahead, and RVs of the MVRVM. ....	32
2.3 Observed versus predicted releases of the MVRVM with 0.90 confidence intervals (shaded region) for the testing phase (2007 irrigation season): (a) Prediction of Sevier Bridge reservoir releases one day ahead, (b) Prediction of Sevier Bridge Reservoir releases two days ahead, (c) Prediction of DMAD reservoir releases one day ahead, (d) Prediction of DMAD reservoir releases two days ahead .....	33
2.4 Bootstrap histogram of the MVRVM model for the RMSE and E test .....	34
2.5 Bootstrap histogram of the number of RVs.....	35
2.6 Observed versus predicted releases of the ANN for the testing phase (2007 irrigation season): (a) Prediction of Sevier Bridge reservoir releases one day ahead, (b) Prediction of Sevier Bridge Reservoir releases two days ahead, (c) Prediction of DMAD reservoir releases one day ahead, (d) Prediction of DMAD reservoir releases two days ahead .....	36
2.7 Bootstrap histogram of the ANN model for the RMSE and E test.....	37
3.1 Lower Sevier River Basin .....	61
3.2 Sevier River, Central Utah Canal, Vincent Canal, and Leamington Canal .....	62
3.3 Observed versus predicted diversions for the hourly MVRVM model with 0.90 confidence intervals (shaded region) for the testing phase (2007 irrigation season).....	63
3.4 Observed versus predicted diversions of the daily MVRVM model with 0.90 confidence intervals (shaded region) for the testing phase (2007 irrigation season).....	64
3.5 Observed versus predicted diversions of the hourly ANN model for the testing phase (2007 irrigation season).....	65

3.6	Observed versus predicted diversions of the daily ANN model for the testing phase (2007 irrigation season).....	66
3.7	Bootstrap histogram of the hourly MVRVM model for the number of RVs. ....	67
3.8	Bootstrap histogram of the daily MVRVM model for the number of RVs. ....	68
3.9	Bootstrap histograms of the hourly MVRVM model for the RMSE test.....	69
3.10	Bootstrap histograms of the daily MVRVM model for the RMSE test.....	70
3.11	Bootstrap histograms of the hourly ANN model for the RMSE test. ....	71
3.12	Bootstrap histograms of the daily ANN model for the RMSE test. ....	72
4.1	Sensor station locations at the study river reach at Lower Sevier River Basin, Utah .....	97
4.2	(a) Daily streamflow on the Sevier River near Lynndyl, (b) Daily water balance in the study river reach .....	98
4.3	Observations versus predictions of the MVRVM with 0.90 confidence intervals (shaded region) for the testing phase (2007 irrigation season): (a) Prediction of streamflow one day ahead, (b) Prediction of streamflow two days ahead, (c) Prediction of net cumulative loss/gain water over one day, (d) Prediction of net cumulative loss/gain water over two days.....	99
4.4	Observations versus predictions of the ANN for the testing phase (2007 irrigation season): (a) Prediction of streamflow one day ahead, (b) Prediction of streamflow two days ahead, (c) Prediction of net cumulative loss/gain water over one day, (d) Prediction of net cumulative loss/gain water over two days .....	100
4.5	Bootstrap histograms of the MVRVM model for the RMSE test. Streamflow predictions: (a) 1-day ahead, (b) 2-days ahead; net cumulative water loss/gain predictions: (c) over one day, (d) over two days .....	101
4.6	Bootstrap histograms of the ANN model for the RMSE test. Streamflow predictions: (a) 1-day ahead, (b) 2-days ahead; net cumulative water loss/gain predictions: (c) over one day, (d) over two days.....	102
4.7	Bootstrap histogram of the MVRVM model for the number of RVs. ....	103
C.1	Plot of observed versus predicted releases of Sevier Bridge reservoir two days ahead, and RVs of the MVRVM. ....	113



C.2	Plot of observed versus predicted releases of DMAD reservoir one day ahead, and RVs of the MVRVM. ....	114
C.3	Plot of observed versus predicted releases of DMAD reservoir two days ahead, and RVs of the MVRVM. ....	115

# CHAPTER 1

## INTRODUCTION

Water use has been increasing world-wide at more than twice the population rate and a number of regions are already facing a water shortage. Irrigated agriculture, responsible for 40% of world food production, accounts for 70% of water withdrawals and over 90% in some developing regions (FAO 2006). Within this context, there is a necessity for more intensive management of this increasingly scarce resource. Therefore, real-time forecasting of the behavior of water resources systems should be given attention so that water managers can have better information about future water availability in order to achieve greater efficiency in water use.

Irrigation system operation can involve interactions between multiple reservoirs, irrigation canals, and the river system. This operation depends on both human behavior and physical processes that are active in the watershed. Human behavior is associated with the reservoir operator who has to release water from multiple reservoirs to fulfill different downstream water requirements. The irrigation canal operator must divert water from the river into a canal taking into consideration water requirements for irrigation, physical characteristics of the irrigation channel system, irrigation scheduling (e.g. rotational or continuous), travel times (from the diversion point to the field outlets), the agricultural area served (e.g. hydrologic, climatic, environmental, etc.) and farmer behavior. An important physical process is the influence of the water loss/gain in a river reach due to transmission loss, tributary contributions, and irrigation return flow to the river. This process is crucial when water loss/gain occurs in the reach between two reservoirs and the reservoir operator needs to manage the quantity of water entering the

reservoir downstream. The combination of all these processes and human decisions may cause unexpected changes in the river system behavior. The behavior of these highly complex systems can become difficult to predict and, more importantly, to generalize. This dissertation seeks to develop forecasting models which have the ability to make accurate multiple predictions of river basin processes and operations and also to generalize well towards future changes in a river basin irrigation system.

Many modeling techniques based on physical principals have been developed to understand the behavior of hydrologic and water resources systems. In physically based modeling, the input-output relationship is obtained by the development and solution of fluid mechanics and thermodynamics equations, with appropriate and detailed boundary conditions, to describe the dynamics of water throughout the hydrologic system in question (Brutsaert 2005). However, physically based solutions are feasible only for some simplified situations; physiographic and geomorphic characteristics of many hydrologic systems are complicated and variable, and have a large degree of uncertainty in the boundary conditions (Brutsaert 2005). In highly complex systems (e.g. a regulated river basin), the practical application of physically based models can be limited by the lack of required data and the expense of data acquisition. To overcome these limitations, researchers have used data-driven models based on machine learning (ML) as an alternative to physically based models (Khalil et al. 2005a; Solomatine and Shrestha 2009). ML theory is related to pattern recognition and statistical inference wherein a model is capable of learning to improve its performance of a task on the basis of its own previous experience (Mjolsness and DeCoste 2001). In the ML approach, a model is

formulated to link the macro-description of the behavior of a system (output) to the behavior of the constituents of this system (inputs) (Guergachi and Boskovic 2008).

Examples of ML models include artificial neural networks (ANNs), support vector machines (SVMs), and relevance vector machines (RVMs). The latter model, the RVM, can be used via its Bayesian approach to avoid overfitting during parameter estimation, to guaranty generalization performance (robustness). ML theory faces the issue of how best to update models on the basis of new data and how to seek parsimony in the model formulation (Mjolsness and DeCoste, 2001). Parsimony is associated with the principal of Ockham's razor which can be translated in ML theory as: "a model should be no more complex than is sufficient to explain the data" (Mjolsness and DeCoste 2001; Tipping 2006). Tipping (2006) stated that the effect of Ockham's razor is an automatic and satisfying consequence of applying the Bayesian framework. In recent years, papers in water resources modeling have demonstrated that applying the RVM approach can result in a parsimonious model capable of a robust prediction of water system state. In addition, they have the capability to estimate the uncertainty of the prediction (Khalil et al. 2005a; Khalil et al. 2005b; Ghosh and Mujumdar 2007). These papers applied RVM models that only allow regression from multivariate inputs to a univariate output variable. This dissertation uses an extension of the RVM model to handle multivariate outputs represented by multiple-time-ahead forecasts for several applications in a regulated river basin system. The proposed model recognizes the patterns between future multiple states of the river basin system as outputs, and past observations collected from this system as inputs.

This research uses the Multivariate Relevance Vector Machine (MVRVM) (Thayananthan et al., 2008) to obtain multivariate predictions with predictive confidence intervals. The MVRVM is a Bayesian regression tool extension of the RVM algorithm developed by Tipping and Faul (2003) to produce multivariate outputs when given a set of multivariate inputs. In addition to its ability to predict multiple outputs, the MVRVM has the same properties as the conventional RVM: high prediction accuracy, robustness, sparse formulation, and characterization of uncertainty in the predictions.

The dissertation addresses the following two hypotheses related to the application of the multivariate Bayesian learning approach in a river basin system:

- Multivariate Bayesian learning machines can be made to work as parsimonious data-driven models for accurate multiple predictions (with confidence intervals) of simultaneous future states of regulated river basin systems.
- Multivariate Bayesian learning machines have the ability to guaranty a robust model which generalizes well when presented with a range of input vectors compared to traditional and widely used machine learning models (e.g. Artificial Neural Networks).

This dissertation consists of five chapters, including this introduction and a summary chapter. Three chapters are used to address the two hypotheses listed above and to describe the MVRVM learning model. A brief description of these chapters follows.

Chapter 2 presents a MVRVM model that simultaneously forecasts water releases one and two days ahead from two reservoirs that are in series. The model is simultaneously applied to the Sevier Bridge (or Yuba) and the Delta-Millard Association Dam (DMAD) Reservoirs located in the Lower Sevier River Basin near Delta, Utah. The

inputs are available past daily data collected by sensors reporting on weather conditions and on the reservoir releases and flows in canals and the river. The results show that the Bayesian learning procedure is capable of producing a very sparse model which learns the input-output patterns with high accuracy consistent with the statistics for the test results. Predictive confidence intervals can also be obtained from the model with this Bayesian approach. The performance results are fairly similar when compared with an ANN. Bootstrap analysis is used to explore the robustness of the MRVM model. Narrow confidence bounds in the bootstrap histograms imply low variability of the test statistics when presented with a range of input vectors, which indicates that the model is robust. The bootstrap histograms show that the MVRVM is more robust than ANNs.

In Chapter 3, two MVRVM models are applied to develop multiple-time-ahead predictions of required hourly and daily diversions for a system of multiple irrigation canals. This model is applied to three irrigation canals, the Central Utah, Vincent, and Leamington Canals, that are located in the Lower Sevier River Basin near Delta, Utah. Inputs for the hourly model are past observations of water diversions for the three canals. The multiple outputs are the predicted diversion demands for the three canals one, twelve, and twenty-four hours ahead. The inputs for the daily model are the past daily observations of water diversions for three canals and climatic data (maximum and minimum daily temperature). The multiple outputs are the predicted diversion demands for the three canals one and two days ahead. Test results show that the MVRVM models generate a sparse formulation and learn the input-output patterns with good prediction accuracy. A bootstrap analysis is used to evaluate robustness of model parameter

estimation. The MVRVM models outperform an artificial neural network (ANN) model in terms of robustness.

Chapter 4 presents a MVRVM model that simultaneously predicts the non-linear behavior of hydrological processes in a river system: streamflow one and two days into the future, and net cumulative water loss/gain in a river reach over the same two-day period. In this chapter we focus on a reach of the Sevier River which is regulated by Sevier Bridge (or Yuba) Reservoir to fulfill downstream irrigation canal orders (Central Utah Canal, Vincent Canal, Leamington Canal) and also provide water to the Delta-Millard Association Dam (DMAD) Reservoir. The inputs are commonly available past data (releases from the upstream reservoir, streamflow, river outflows and air temperature). The results show that the model learns the patterns between outputs and inputs for the training phase and makes accurate predictions for the testing phase. The performance results for the MVRVM and ANN models are very similar in terms of prediction efficiency. However, a bootstrap analysis shows that the MVRVM achieves more robust performance.

## References

- Brutsaert W (2005) Hydrology, an introduction. Cambridge University Press, NY.
- FAO. 2006. Water monitoring-mapping existing global system and initiatives. Prepared by FAO on behalf of the UN-Water Task Force on Monitoring. [http://www.fao.org/nr/water/docs/UNW\\_MONITORING\\_REPORT.pdf](http://www.fao.org/nr/water/docs/UNW_MONITORING_REPORT.pdf). Accessed 15 January 2010.
- Ghosh S, Mujumdar PP (2007) Statistical downscaling of GCM simulations to streamflow using relevance vector machine, *Adv Water Resour* 31:132-146.
- Guergachi A, Boskovic G (2008) System models or learning machines? *Appl Math Comp* 204:553–567.

- Khalil A, McKee M, Kemblowski MW, Asefa T (2005a) Sparse Bayesian learning machine for real-time management of reservoir releases. *Water Resour Res* 41:W11401.
- Khalil A, McKee M, Kemblowski MW, Asefa T, Bastidas L (2005b) Multiobjective analysis of chaotic dynamic systems with sparse learning machines. *Adv Water Resour* 29:72-88.
- Mjolsness E, DeCoste D (2001) Machine learning for science: state of the art and future prospects. *Science* 293:2051–2055.
- Solomatine DP, Shrestha DL (2009) A novel method to estimate model uncertainty using machine learning techniques, *Water Resour Res* 45:W00B11.
- Thayananthan A, Navaratnam R, Stenger B, Torr PHS, Cipolla R (2008) Pose estimation and tracking using multivariate regression. *Pattern Recognit Lett* 29(9):1302-1310.
- Tipping M (2006) Bayesian inference: an introduction to principles and practice in machine learning. *Adv Lect Mach Learn* 41-62.
- Tipping M, Faul A (2003) Fast marginal likelihood maximization for sparse Bayesian models, paper presented at Ninth International Workshop on Artificial Intelligence and Statistics, Soc. for Artif Intel Stat, Key West, FL.



## CHAPTER 2

### MULTIVARIATE BAYESIAN REGRESSION APPROACH TO FORECAST RELEASES FROM A SYSTEM OF MULTIPLE RESERVOIRS<sup>1</sup>

#### **Abstract**

This research presents a model that simultaneously forecasts water releases one and two days ahead from two reservoirs in series. In practice, multiple reservoir system operation is a difficult process that involves many decisions for real-time water resources management. The operator of the reservoirs has to release water from more than one reservoir taking into consideration different water requirements (irrigation, environmental issues, hydropower, recreation, etc.) in a timely manner. A model that forecasts the required real-time releases in advance from a multiple reservoir system could be an important tool to allow the operator of the reservoir system to make better-informed decisions for releases needed downstream. The model is developed in the form of a multivariate relevance vector machine (MVRVM) that is based on a sparse Bayesian regression model approach. With this Bayesian approach, a predictive confidence interval is obtained from the model that captures the uncertainty of both the model and the data. The model is applied to the multiple reservoir system located in the Lower Sevier River Basin near Delta, Utah. The results show that the model learns the input-output patterns with high accuracy. Computing multiple-time-ahead predictions in real-time would require a model which guarantees not only good prediction accuracy but also robustness with respect to future changes in the nature of the inputs data. A bootstrap analysis is used to guarantee good generalization ability and robustness of the MVRVM. Test results

---

<sup>1</sup> Coauthored by Andres M Ticlavilca and Mac McKee

demonstrate good performance of predictions and statistics that indicate robust model generalization abilities. The MVRVM is compared in terms of performance and robustness with another multiple output model such as Artificial Neural Network (ANN).

## **2.1 Introduction**

The per capita availability of water resources is decreasing world-wide due to population growth, climate change, rapidly increasing demands (for irrigation, domestic supply, recreation, etc.), and pollution. In order to achieve greater operational efficiency in meeting these increasing demands and decreasing relative supplies, water managers must have better information about future conditions of their water systems. The purpose of this paper is to use a real-time model that can provide valuable information to the operator of a multiple reservoir system in the form of multi-time-ahead release predictions with predictive confidence intervals.

Techniques based on physical modeling have been developed to characterize current and future states of water resources systems. Their practical applications are often limited by the lack of required data and the expense of data acquisition (Khalil et al. 2005a). To overcome these limitations, researchers have used data-driven modeling as an alternative to physically based models (Lobrecht and Solomatine 2002; Khalil et al. 2005a). Examples of such models include artificial neural network (ANNs), support vector machines (SVMs) and relevance vector machines (RVMs). These types of models are derived from the emerging area of machine-learning theory. They are characterized by their ability to capture the underlying physics of the system simply by examination of the inputs and outputs of the system. They can be used to provide predictions of the system behavior using only historical data; they “let the data speak” (Khalil et al. 2005b).

Dams and reservoir systems have been built to regulate storage and manage water distribution. However, many of these systems are not producing benefits that would economically justify their development (World Commission on Dams 2000; Labadie 2004). As a result, we must focus on improving the operational effectiveness of existing reservoir systems to maximize their value (Labadie 2004). Providing a model that forecasts required releases from a system of multiple reservoirs could be an important tool in integrated reservoir operation and management. Forecasts of required releases can allow the operator of the multiple reservoir system to make better-informed decisions for releases needed downstream.

The multiple reservoir system operation depends on physical behavior of the watershed (hydrologic, climatic, environmental, etc.) and human behavior. Human behavior takes the form of the reservoir operator who has to release water from more than one reservoir to fulfill different water requirements. The combination of all these behaviors may cause unexpected future changes in the reservoir system operation which could be extremely difficult to predict. Therefore, it is necessary to develop predictive models which have the ability to guarantee robustness towards future changes in the system behavior.

Khalil et al. (2005b) applied RVM, which is based on Bayesian learning theory, to predict the real-time operation of a single reservoir. The target output of their model is the hourly prediction of the quantity of water to be released from a single reservoir in order to meet downstream diversion requirements. Their performance results showed that the RVM model was able to predict future system states (generalization ability) and had the capability to estimate the uncertainty of the predictions (predictive confidence

intervals). The research reported here extends this capability to a multiple-day-ahead forecast for a multi-reservoir setting.

In order to obtain multiple-time-ahead predictions with (predictive) confidence intervals, the model exploits the capability of the Multivariate Relevance Vector Machine (MVRVM) (Thayananthan 2005). The model forecasts the water releases of two reservoirs simultaneously having as inputs recent historical data on reservoir releases, diversions into canals, weather, and streamflows. The target outputs are the predictions of the required water releases from two reservoirs. These predictions are made one and two days ahead, simultaneously for each reservoir. Therefore, the model recognizes the patterns between future reservoir releases and historical data collected from the system.

The MVRVM is a Bayesian regression tool extension of the RVM algorithm developed by Tipping and Faul (2003) to produce multivariate outputs when given a set of inputs. In addition to its ability to predict multiple outputs, the MVRVM has the same properties of the conventional RVM: high prediction accuracy, robustness and characterization of uncertainty in the predictions. Therefore, developing a model with all these properties can work as a practical decision support tool in real-time water resources management by providing multiple predictions that are difficult (or not practical) to obtain from traditional modeling approaches.

The remainder of the paper describes the MVRVM learning model, the area of study where the model has been applied, how the model has been developed for a multiple reservoir system, the results of the MVRVM application, the comparison with the performance of an ANN model, and conclusions that can be drawn.

## 2.2 Model Description

Thayananthan (2005) proposed the Multivariate Relevance Vector Machine (MVRVM) to provide a regression tool capable of generating multivariate outputs. This model is an extension of the sparse Bayesian model developed by Tipping and Faul (2003). It is developed as follows.

Given a training data set of input-target vector pairs  $\{\mathbf{x}^{(n)}, \mathbf{t}^{(n)}\}_{n=1}^N$ , where  $N$  is the number of observations,  $\mathbf{x} \in \mathbb{R}^D$  is a  $D$ -dimensional input vector,  $\mathbf{t} \in \mathbb{R}^M$  is a  $M$ -dimensional output target vector, the model learns the dependency between input and output target with the purpose of making accurate predictions of  $\mathbf{t}$  for previously unseen values of  $\mathbf{x}$ :

$$\mathbf{t} = \mathbf{W} \Phi(\mathbf{x}) + \boldsymbol{\varepsilon} \quad (2.1)$$

where  $\mathbf{W}$  is a  $M \times P$  weight matrix and  $P = N+1$ . A fixed kernel function  $K(\mathbf{x}, \mathbf{x}^{(n)})$  (Appendix A) is used to create a vector of basis functions of the form  $\Phi(\mathbf{x}) = [1, K(\mathbf{x}, \mathbf{x}^{(1)}), \dots, K(\mathbf{x}, \mathbf{x}^{(N)})]$ . The error  $\boldsymbol{\varepsilon}$  is conventionally assumed to be zero-mean Gaussian with diagonal covariance matrix  $\mathbf{S} = \text{diag}(\sigma_1^2, \dots, \sigma_M^2)$ .

Let  $\mathbf{t} = [\tau_1, \dots, \tau_r, \dots, \tau_M]^T$  and  $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_r, \dots, \mathbf{w}_M]^T$ . A likelihood distribution of the weight matrix can be written as a product of Gaussians of the weight vectors ( $\mathbf{w}_r$ ) corresponding to each target output ( $\tau_r$ ) (Thayananthan et al. 2008):

$$p(\{\mathbf{t}^{(n)}\}_{n=1}^N | \mathbf{W}, \mathbf{S}) = \prod_{n=1}^N N(\mathbf{t}^{(n)} | \mathbf{W} \Phi(\mathbf{x}^{(n)}), \mathbf{S}) = \prod_{r=1}^M N(\tau_r | \mathbf{w}_r \Phi, \sigma_r^2) \quad (2.2)$$

where  $\Phi = [1, \Phi(\mathbf{x}_1), \Phi(\mathbf{x}_2), \dots, \Phi(\mathbf{x}_N)]$ . Eq. 2.2 therefore contains the parameters  $\mathbf{W}$  (which is  $M$  by  $N$ ) and  $\sigma_r^2$  (which is a vector of length  $M$ ). As a result, there is a danger that the maximum likelihood estimation of  $\mathbf{w}_r$  and  $\sigma_r^2$  will suffer from severe over-fitting.

To avoid this, Tipping (2001) proposed constraining the selection of parameters by applying a Bayesian perspective and defining an explicit zero-mean Gaussian prior probability distribution over them (Thayananthan et al. 2008):

$$p(\mathbf{W} | \mathbf{A}) = \prod_{r=1}^M \prod_{j=1}^P N(w_{rj} | 0, \alpha_j^{-2}) = \prod_{r=1}^M N(\mathbf{w}_r | 0, \mathbf{A}) \quad (2.3)$$

where  $\mathbf{A} = \text{diag}(\alpha_1^{-2}, \dots, \alpha_P^{-2})^T$  is a diagonal matrix of hyperparameters  $\alpha_j$ , and  $w_{rj}$  is the  $(r,j)^{\text{th}}$  element of the weight matrix  $\mathbf{W}$ . Each  $\alpha_j$  controls the strength of the prior over its associated weight (Tipping and Faul 2003).

Bayesian inference considers the posterior distribution of the model parameters, which is proportional to the product of the likelihood and prior distributions:

$$p(\mathbf{W} | \{\mathbf{t}\}_{n=1}^N, \mathbf{S}, \mathbf{A}) \propto p(\{\mathbf{t}\}_{n=1}^N | \mathbf{W}, \mathbf{S}) p(\mathbf{W} | \mathbf{A}) \quad (2.4)$$

The posterior parameter distribution conditioned on the data can be written as the product of Gaussians for the weight vectors of each target output (Thayananthan et al. 2008):

$$p(\mathbf{W} | \{\mathbf{t}\}_{n=1}^N, \mathbf{S}, \mathbf{A}) \propto p(\{\mathbf{t}\}_{n=1}^N | \mathbf{W}, \mathbf{S}) p(\mathbf{W} | \mathbf{A}) \propto \prod_{r=1}^M N(\mathbf{w}_r | \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r) \quad (2.5)$$

The posterior distribution of the weights is Gaussian  $N(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r)$  with the covariance and mean,  $\boldsymbol{\Sigma}_r = (\mathbf{A} + \sigma_r^{-2} \boldsymbol{\Phi}^T \boldsymbol{\Phi})^{-1}$  and  $\boldsymbol{\mu}_r = \sigma_r^{-2} \boldsymbol{\Sigma}_r \boldsymbol{\Phi}^T \boldsymbol{\tau}_r$ , respectively. Given this posterior, we can obtain an optimal weight matrix by estimating a set of hyperparameters that maximizes the data likelihood over the weights in Equation (2.5) (Thayananthan et al. 2008). The marginal likelihood is then:

$$\begin{aligned} p(\{\mathbf{t}\}_{n=1}^N | \mathbf{A}, \mathbf{S}) &= \int p(\{\mathbf{t}\}_{n=1}^N | \mathbf{W}, \mathbf{S}) p(\mathbf{W} | \mathbf{A}) d\mathbf{W}, \\ &= \prod_{r=1}^M \int N(\boldsymbol{\tau}_r | \mathbf{w}_r \boldsymbol{\Phi}, \sigma_r^2) N(\mathbf{w}_r | \mathbf{0}, \mathbf{A}) d\mathbf{w}_r = \prod_{r=1}^M |\mathbf{H}_r|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \boldsymbol{\tau}_r^T \mathbf{H}_r^{-1} \boldsymbol{\tau}_r\right) \end{aligned} \quad (2.6)$$

where  $\mathbf{H}_r = \sigma_r^2 \mathbf{I} + \Phi \mathbf{A}^{-1} \Phi^T$ . Then, we can obtain an optimal set of hyperparameters  $\boldsymbol{\alpha}^{\text{opt}} = \{\boldsymbol{\alpha}_j^{\text{opt}}\}_{j=1}^P$  and noise parameters  $(\boldsymbol{\sigma}^{\text{opt}})^2 = \{\sigma_r^{\text{opt}}\}_{r=1}^M$  by maximizing the marginal likelihood using the fast marginal likelihood maximization algorithm proposed by Tipping and Faul (2003). During the optimization process, many elements of  $\boldsymbol{\alpha}$  go to infinity, for which the corresponding posterior probability of the weight becomes zero. The relatively few nonzero weights correspond to the input vectors that form the sparse core of the RVM model. It is these input vectors, called the relevance vectors (RVs), that generate a sparse representation. Inducing sparsity can be an effective method to control model complexity, avoid over-fitting and control computational characteristics of model performance (Tipping and Faul 2003). The optimal parameters are used to obtain the optimal weight matrix with optimal covariance  $\boldsymbol{\Sigma}^{\text{opt}} = \{\boldsymbol{\Sigma}_r^{\text{opt}}\}_{r=1}^M$  and mean  $\boldsymbol{\mu}^{\text{opt}} = \{\boldsymbol{\mu}_r^{\text{opt}}\}_{r=1}^M$ .

We can compute the predictive distribution for any new input  $\mathbf{x}^*$ , corresponding to a target  $\mathbf{t}^*$  (Tipping 2001), from:

$$p(\mathbf{t}^* | \mathbf{t}, \boldsymbol{\alpha}^{\text{opt}}, (\boldsymbol{\sigma}^{\text{opt}})^2) = \int p(\mathbf{t}^* | \mathbf{W}, (\boldsymbol{\sigma}^{\text{opt}})^2) p(\mathbf{W} | \mathbf{t}, \boldsymbol{\alpha}^{\text{opt}}, (\boldsymbol{\sigma}^{\text{opt}})^2) d\mathbf{W} \quad (2.7)$$

Taking into consideration that both terms in the integrand are Gaussian, Equation (2.7) is computed as:

$$p(\mathbf{t}^* | \mathbf{t}, \boldsymbol{\alpha}^{\text{opt}}, (\boldsymbol{\sigma}^{\text{opt}})^2) = \mathcal{N}(\mathbf{t}^* | \mathbf{y}^*, (\boldsymbol{\sigma}^*)^2) \quad (2.8)$$

where  $\mathbf{y}^* = [y_1^*, \dots, y_r^*, \dots, y_M^*]^T$  is the predictive mean with  $\mathbf{y}_r^* = (\boldsymbol{\mu}_r^{\text{opt}})^T \Phi(\mathbf{x}^*)$ ; and  $(\boldsymbol{\sigma}^*)^2 = [(\sigma_1^*)^2, \dots, (\sigma_r^*)^2, \dots, (\sigma_M^*)^2]^T$  is the predictive variance with  $(\sigma_r^*)^2 = (\sigma_r^{\text{opt}})^2 + \Phi(\mathbf{x}^*)^T \boldsymbol{\Sigma}_r^{\text{opt}} \Phi(\mathbf{x}^*)$ .  $(\sigma_r^*)^2$  contains the sum of two variance terms: the noise on the data and the uncertainty in the prediction of the weight parameters (Tipping 2001).

The standard deviation  $\sigma_r^*$  of the predictive distribution is defined as a predictive error bar of  $y_r^*$  (Bishop 1995). Then, the width of the 90% predictive confidence interval for any  $y_r^*$  can be calculated as  $\pm 1.65 \sigma_r^*$ .

Readers interested in greater detail regarding multivariate sparse Bayesian regression, its mathematical formulation, and the optimization procedures of the model are referred to Thayananthan (2005), Thayananthan et al. (2008), Tipping (2001), and Tipping and Faul (2003). A MATLAB code developed by Thayananthan (2005) is available from <http://mi.eng.cam.ac.uk/~at315/MVRVM>.

### **2.3 Study Area**

The Sevier River Basin is used here to demonstrate the MVRVM modeling approach as applied to the operation of multiple reservoirs. The basin located in south-central Utah and is the largest area drainage in the state (approximately 12.5 percent of the state's area). Average annual precipitation ranges from 6.4 to 13.0 inches in the valleys to more than 40 inches in the high mountains. Elevation, precipitation, and temperatures are highly variable over the basin, and as a result there are several vegetative types that grow in the area. The population of the basin was more than 56,700 people in 1997, with most of it residing in small farming communities. The population of the basin is expected to reach over 86,000 people by 2020 (based on the current annual growth rate of 1.82%). The economy of the Basin is based primarily on agriculture and also there are other important economic activities such as tourism, few mining and manufacturing enterprises (Berger et al. 2003).

The Sevier River Basin is a highly instrumented and controlled basin. Automated data collection equipment (sensors on reservoirs, canals, diversions, and the river itself)



was installed beginning in 1999 (Berger et al. 2002), and is currently collecting data at stations throughout the entire basin. The data collected are displayed in the Sevier River Water Users website, [www.sevierriver.org](http://www.sevierriver.org). This website provides the user several data retrieval and display options, such as hourly flow data for the previous 7 days or the current river and canal flow information displayed in spatial diagrams (Berger et al. 2002).

In this paper we focus on the Lower Sevier River Basin, which is regulated by three reservoirs: Sevier Bridge (or Yuba) Reservoir, the Delta-Millard Association Dam (DMAD) Reservoir, and Gunnison Bend Reservoir (Fig. 2.1).

#### **2.4 Model Application to Multiple Reservoir System**

The MVRVM previously described is applied to the multiple reservoir system located in the Lower Sevier River Basin. Monitoring data are posted hourly to the Sevier River website. These data enable real-time operations of reservoir releases and canal diversions. Daily average flow, reservoir release, and canal diversion data taken from the Sevier River database from 2001 to 2007 were used to build the MVRVM reservoir model. Daily data from the irrigation seasons of 2001 through 2006 were used to train the MVRVM and find the model parameters. Daily data from the 2007 irrigation season were used to test the model.

The inputs are available historic daily data collected by sensors on the reservoirs, canals, diversions, weather and the river itself. The multiple output target vectors are the predictions of required water releases one and two days ahead from Sevier Bridge and DMAD reservoirs.

The gauging station located on the Sevier River near Juab measures the releases from Sevier Bridge Reservoir into the Sevier River, and the station located on the Sevier River below DMAD Reservoir measures the releases from DMAD reservoir into the Sevier River (Fig. 2.1).

Four sensor stations are located between Sevier Bridge and DMAD reservoirs: three measure diversions to irrigation canals, and one station measures streamflow on the Sevier River near Lynndyl. One major diversion is made from DMAD reservoir to an irrigation canal labeled Canal A in Figure 2.1. Three diversions from Gunnison Bend reservoir are measured, corresponding to irrigation releases to serve the Abraham Canal, the Deseret High Canal, and the Deseret Low Canal.

Daily maximum and minimum temperature from Delta City were obtained from the Community Environmental Monitoring Program (CEMP) website, and daily maximum and minimum temperature from Oak City were obtained from the National Oceanic and Atmospheric Administration (NOAA).

Inputs to the MVRVM model also include past data on daily Sevier Bridge reservoir releases and data for DMAD Reservoir releases.

The inputs used in the model to predict reservoir releases are expressed as:

$$\mathbf{x} = [\mathbf{X1}_{d-nd}, \mathbf{X2}_{d-nd}, \mathbf{X3}_{d-nd}, \mathbf{X4}_{d-nd}, \mathbf{X5}_{d-nd}]^T \quad (2.9)$$

where,

d= day of prediction

nd= number of days previous to the prediction time

$\mathbf{X1}_{d-nd}$  = diversions to the Central Utah canal, Vincent canal, Leamington canal, canal A, Abraham canal, Deseret High canal, and Deseret Low canal.

$\mathbf{X2}_{d-nd}$  = streamflow on the Sevier River near Lynndyl.

$\mathbf{X3}_{d-nd}$  = Sevier Bridge reservoir releases.

$\mathbf{X4}_{d-nd}$  = DMAD reservoir releases.

$\mathbf{X5}_{d-nd}$  = maximum and minimum daily temperature from Oak City and Delta City.

The multiple output target vector of the model is expressed as:

$$\mathbf{t} = [\mathbf{R1}_d, \mathbf{R1}_{d+1}, \mathbf{R2}_d, \mathbf{R2}_{d+1}]^T \quad (2.10)$$

where,

$\mathbf{R1}_d$  = prediction of Sevier Bridge reservoir release one day ahead

$\mathbf{R1}_{d+1}$  = prediction of Sevier Bridge reservoir release two days ahead

$\mathbf{R2}_d$  = prediction of DMAD reservoir releases one day ahead

$\mathbf{R2}_{d+1}$  = prediction of DMAD reservoir releases two day ahead

Finally, the model can be defined as in Eq. 2.1 with a data set of input-output pairs

$\{\mathbf{x}^{(n)}, \mathbf{t}^{(n)}\}_{n=1}^N$ , where N is the number of observations.

## 2.5 Results and Discussion

### 2.5.1 Model selection

In Eq. 1, the basis function ( $\Phi$ ) is defined in terms of a fixed kernel function. It is necessary to choose the type of kernel function and also to determine the values for its associated parameter, the kernel width (Tipping 2001). The statistics used for the selection of the model are the coefficient of efficiency (E) and the correlation coefficient (r). E is equal to 1 minus the ratio of the mean square error to the variance in the observed data (Appendix B). This statistic ranges from minus infinity (poor model) to 1.0 (a

perfect model) (Legates and McCabe 1999). The  $r$  value measures the correlation between observed and predicted reservoir release.

Several MVRVM models were built with variation in the type of kernel, kernel width and the number of days previous to the prediction time (from 1 to 5 days). The model selected was the one with the maximum  $E$  of the average outputs corresponding to the testing phase. Table 2.1 shows the selected model for each type of kernel. The model with Gaussian kernel shows the highest  $E$  and  $r$  of the average results. Therefore it was selected as the best type of kernel function that describes the input-output patterns for the model. From Table 2.1, we can see also that the selected model requires two prior days of historical data as input.

### 2.5.2 Performance evaluation

Fig. 2.2 illustrates the training phase of Sevier Bridge reservoir release prediction one day ahead. The training phase of release predictions for Sevier Bridge reservoir two days ahead and DMAD reservoir one and two days ahead are shown in Appendix C.

The relevance vectors (RVs) are subsets of the training data set that are used for prediction (Khalil et al. 2005b); the complexity of the model is proportional to the number of RVs. The model only utilizes 39 RVs from the full data set (1248 observations) that was used for training (2001 through 2006 irrigation seasons). This small number of vectors illustrates that the Bayesian learning procedure embodied in the MVRVM is capable of producing very sparse models.

The RVs are the most essential features (observations) of the training data set around which the MVRVM is built (Khalil et al. 2005b). The MVRVM identifies the greatest number of RVs (10 RVs) from the 2006 irrigation season (Fig. 2.2f), and the

lowest number (one RV) from the 2004 season (Fig. 2.2d). The 2006 irrigation season has the largest number of relevance observations, while the majority of the observations from the 2004 irrigation season have been ignored to build the model.

The predicted outputs of the MVRVM for the testing phase (2007 irrigation season) are shown as the full lines in Fig. 2.3. The figure shows good performance of the machine. The model explains well the observed releases (dots) for the releases one day ahead for both reservoirs. The releases two days ahead from Sevier Bridge Reservoir also illustrated good performance (Fig. 2.3a, Fig. 2.3b, and Fig. 2.3c). The performance accuracy is reduced for DMAD Reservoir releases two days ahead (Fig. 2.3d). This accuracy reduction is found in most of the multiple-time-ahead prediction models, where the farther we predict into the future, the less accurate the prediction becomes.

Fig. 2.3 also shows the 0.90 confidence interval (shaded region) associated with the predictive variance of the MVRVM in Eq. 8. The confidence intervals for the two-day-ahead prediction (Fig. 2.3b and Fig. 2.3d) become wider than the confidence interval for their corresponding one-day-ahead predictions for both reservoirs (Fig. 2.3a and Fig. 2.3c). We can see how the uncertainty in the predictions increases when predicting further into the future.

Table 2.2 shows statistical measures of MVRVM performance for both the training and testing phases. Again, we can see good performance of the machine in the testing phase for the one-day-ahead prediction for both reservoir releases ( $R1_d$  and  $R2_d$ ) with a high E, 0.95 and 0.93, respectively, for Sevier Bridge and DMAD reservoirs; and the two-day-ahead prediction for Sevier Bridge Reservoir releases ( $R1_{d+1}$ ) also has a high

E, 0.87. The performance accuracy is reduced for the two-day-ahead prediction of DMAD Reservoir releases ( $R_{d+1}$ ) with a lower E of 0.80.

### 2.5.3 Bootstrap analysis

It is necessary to develop models to guarantee not only high accuracy but also good generalization and robustness of model parameter estimation with respect to future changes in the nature of the input data. Changes in the training data used to build a model may give different test results. Different sets of training data may produce models with very different generalization accuracies. The bootstrap method was used to explore the implications of the change in the nature of input data and to guarantee good generalization ability and robustness of the MVRVM (Khalil et al. 2005b).

The bootstrap data set was created by randomly selecting from the whole training data set, with replacement. Because the selection is from the whole training data set, there is nearly always duplication of individual points in a bootstrap data set. In the bootstrap estimation, this selection process was independently repeated 1,000 times to yield 1,000 bootstrap training data sets, which are treated as independent sets (Duda et al. 2001). For each of the bootstrap training data sets, a model was built and evaluated over the original test data set. Fig. 2.4 shows the bootstrap histograms based on 1,000 bootstrap training data sets of the E and RMSE test.

Efron and Tibshirani (1998) emphasized that it is always wise to look at the bootstrap data graphically, rather than relying entirely on a single summary statistic estimator. The bootstrap method provides information on the uncertainty in the statistics estimator evaluated in the model. The width of the bootstrapping confidence intervals provides information on the uncertainty in the model parameters. A narrow confidence

interval implies low variability of the statistics with respect to possible future changes in the nature of the input data, which indicates that the model is robust (Khalil et al. 2005b). According to Khalil et al. (2005b) a robust model is one that shows narrow confidence bounds in the bootstrap histograms, such as those illustrated in Fig. 2.4.

Fig. 2.5 shows narrow confidence bounds on the bootstrap histogram of the number of RVs used in the MVRVM model. The low variability of the number of RVs indicates that the model structure is stable and robust with respect to future changes in the nature of the input data.

#### 2.5.4 Comparison between MVRVM and ANN

ANNs have been widely applied in hydrology and water resources modeling (ASCE Task Committee on the Application of ANNs in Hydrology 2000a, b; Khalil et al. 2005c; Adeloye 2009). A comparative analysis between the MVRVM developed here and ANNs is performed in terms of performance and robustness.

The ANN toolbox available in MATLAB is applied in this research. There are many functions to train the ANN and find the model parameters. Conjugate gradient training functions (i.e. conjugate gradient with Powell-Beale restarts, and scale conjugated gradient) adjust the ANN parameters faster than basic training algorithms. The Quasi-Newton training function optimizes the ANN parameters by using Newton's method with an approximate Hessian matrix for fast optimization. The Levenberg-Marquard training function is a fast algorithm with second-order convergence to optimize the ANN parameters. Readers interested in greater detail regarding ANNs and their training functions are referred to Demuth et al. (2009).

Several feed-forward ANN models were trained and tested with variation in the type of training function, size of hidden layer, and the number of days of historic data previous to the prediction time (from 1 to 5 days) required as input. The model selected was the one with the maximum E of the average outputs corresponding to the testing phase.

Table 2.3 shows the selected model for each type of training function. The ANN model that was developed using the scaled-conjugated-gradient-training function (Demuth et al. 2009) shows the highest E and R of the average results. Therefore it was selected as the best type of training function that describes the input-output patterns for the ANN model.

The observed (dots) and predicted (full lines) outputs of the ANN for the testing phase (2007 irrigation season) are shown in Fig. 2.6. Table 2.4 shows the ANN performance for both the training and testing phases. From Table 2.4 we can see that the performance results are fairly similar to the MVRVM performance (Table 2.2).

Fig. 2.7 shows the bootstrap histograms based on 1,000 bootstrap training data sets of the ANN model for the RMSE and E test. The bootstrapped histograms of the MVRVM model (Fig. 2.4) show very narrow confidence bounds in comparison to the histograms of the ANN model (Fig. 2.7). Therefore, the MVRVM appears to be more robust.

## **2.6 Summary and Conclusions**

This paper presents a first attempt to develop multiple-time-ahead predictions of required daily releases from a multiple reservoir system using a MVRVM model. The model is illustrated by application to the Lower Sevier River near Delta, Utah. The



predictions are required water releases one and two days into the future from Sevier Bridge and DMAD reservoirs.

The results show that the model learns the input-output patterns with high accuracy consistent with the statistics for the test results. The statistical results indicate good performance of the model for the one-day prediction for the releases of Sevier Bridge and DMAD reservoirs. The performance decreased slightly for the two-day prediction of DMAD reservoir release.

The MVRVM model has the property of sparse formulation. The model only utilizes 39 RVs from the full data set (out of a possible 1248 observations that were used for training). The parsimonious structure of this empirical model is sufficient to explain the data and to avoid data over-fitting. Therefore, we can see an important advantage of the Bayesian learning procedure, which is the capability of the MVRVM to produce very sparse models.

Another important advantage of utilizing MVRVM is its generalization capabilities while achieving sparse representation. Generalization ability is associated with the capability of the model to predict future system states when presented with a range of input vectors. Multiple reservoir system operation could become a difficult process to predict since this involves many decisions for real-time water resources management. The model presented here ensures good generalization providing robustness with respect to new input data.

The performance results are fairly similar for both the MVRVM and ANN models. However, the bootstrap histograms of the MVRVM model show narrower

confidence bounds in comparison to the histograms of the ANN model. Therefore, the MVRVM is more robust.

In summary, the results presented in this paper have demonstrated the successful performance and robustness of MVRVM for multiple reservoir release forecasts. Simultaneous multiple-time-ahead release predictions from a multiple reservoir system have potential value to assist the reservoir operator in efficiently selecting the real-time operation and management decisions for available water resources.

## References

- Adeloye A J (2009) Multiple linear regression and artificial neural networks models for generalized reservoir storage-yield-reliability function for reservoir planning. *J Hydrol Eng* 14(7):731-738.
- ASCE Task Committee on the Application of ANNs in Hydrology (2000a) Artificial neural networks in hydrology, I: preliminary concepts. *J Hydrol Eng* 5(2):115-123.
- ASCE Task Committee on the Application of ANNs in Hydrology (2000b) Artificial neural networks in hydrology, II: hydrologic application. *J Hydrol Eng* 5(2):124-137.
- Berger B, Hansen R, Hilton A (2002) Using the world-wide-web as a support system to enhance water management, paper presented at the 18<sup>th</sup> ICID Congress and 53<sup>rd</sup> IEC Meeting, Int Comm on Irrig and Drain, Montreal, Quebec, Canada.
- Berger B, Hansen R, Jensen R (2003) Sevier river basin system description, Sevier River Water Users Association, Delta, Utah.
- Bishop CM (1995) *Neural network for pattern recognition*. Oxford University Press, NY.
- Demuth H, Beale M, Hagan M (2009) *Neural network toolbox user's guide*, The MathWorks Inc, MA, USA.
- Duda RO, Hart P, Stork D (2001) *Pattern classification*, Second Edition, edited by Wiley Interscience, NY.
- Efron B, Tibshirani R (1998) *An introduction of the bootstrap*, Monographs on Statistics and Applied Probability 57, CRC Press LLC, Boca Raton, FL.

- Khalil A, McKee M, Kemblowski MW, Asefa T (2005a) Sparse Bayesian learning machine for real-time management of reservoir releases. *Water Resour Res* 41:W11401.
- Khalil A, McKee M, Kemblowski MW, Asefa T, Bastidas L (2005b) Multiobjective analysis of chaotic dynamic systems with sparse learning machines. *Adv Water Resour* 29:72-88.
- Khalil A, McKee M, Kemblowski M, Asefa T (2005c) Basin-scale water management and forecasting using neural networks. *J Am Water Resour Res* 41(1):195-208.
- Labadie JW (2004) Optimal operation of multireservoir systems: state-of-the-art review. *J Water Resour Plann Manage* 130(2):93-111.
- Legates DR, McCabe GJ (1999) Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. *Water Resour Res* 35(1):233-241.
- Lobrecht A H, Solomatine DP (2002) Machine learning in real-time control of water systems. *Urban Water* 4:283-289.
- Thayananthan A (2005) Template-based pose estimation and tracking of 3D hand motion. PhD thesis, Department of Engineering, University of Cambridge, Cambridge, United Kingdom.
- Thayananthan A, Navaratnam R, Stenger B, Torr PHS, Cipolla R (2008) Pose estimation and tracking using multivariate regression. *Pattern Recognit Lett* 29(9):1302-1310.
- Tipping ME (2001) Sparse Bayesian learning and the relevance vector machine. *J Mach Learn* 1:211–244.
- Tipping M, Faul A (2003) Fast marginal likelihood maximization for sparse Bayesian models, paper presented at Ninth International Workshop on Artificial Intelligence and Statistics, Soc. for Artif Intel Stat, Key West, FL.
- World Commission on Dams (2000) Dams and development: a new framework for decision-making. Earthscan Publications Ltd, London and Sterling, VA.

**Table 2.1** Selected MVRVM for each type of kernel function

Testing phase								
Type of kernel	Kernel	Number of						
function	width	days <sup>1</sup>	Statistics	R1 <sub>d</sub>	R1 <sub>d+1</sub>	R2 <sub>d</sub>	R2 <sub>d+1</sub>	Average
<b>Gauss</b>	<b>2900</b>	<b>2</b>	<b>E</b>	0.947	0.868	0.930	0.805	0.888
			<b>R</b>	0.973	0.932	0.968	0.909	0.946
<b>Laplace</b>	<b>3100</b>	<b>1</b>	<b>E</b>	0.925	0.841	0.900	0.780	0.861
			<b>R</b>	0.964	0.917	0.952	0.892	0.931
<b>Cauchy</b>	<b>2900</b>	<b>2</b>	<b>E</b>	0.943	0.866	0.935	0.800	0.886
			<b>R</b>	0.972	0.931	0.969	0.903	0.944
<b>Cubic</b>	<b>1700</b>	<b>2</b>	<b>E</b>	0.936	0.842	0.927	0.813	0.879
			<b>R</b>	0.968	0.918	0.966	0.914	0.942

<sup>1</sup> Number of prior days of historical data required as input

**Table 2.2** MVRVM performance using different statistics

Statistics	Multivariate Relevance Vector Machine							
	Training				Testing			
	$R1_d$	$R1_{d+1}$	$R2_d$	$R2_{d+1}$	$R1_d$	$R1_{d+1}$	$R2_d$	$R2_{d+1}$
<b>Coefficient of efficiency E</b>	0.95	0.87	0.89	0.77	0.95	0.87	0.93	0.80
<b>Correlation coefficient R</b>	0.97	0.93	0.94	0.88	0.97	0.93	0.97	0.91
<b>Root mean square error RMSE, cfs</b>	64.42	99.24	34.84	49.66	59.45	93.82	32.55	54.49
<b>Predictive error, cfs</b>	66.39	101.76	35.82	51.00	66.40	101.78	35.82	51.00

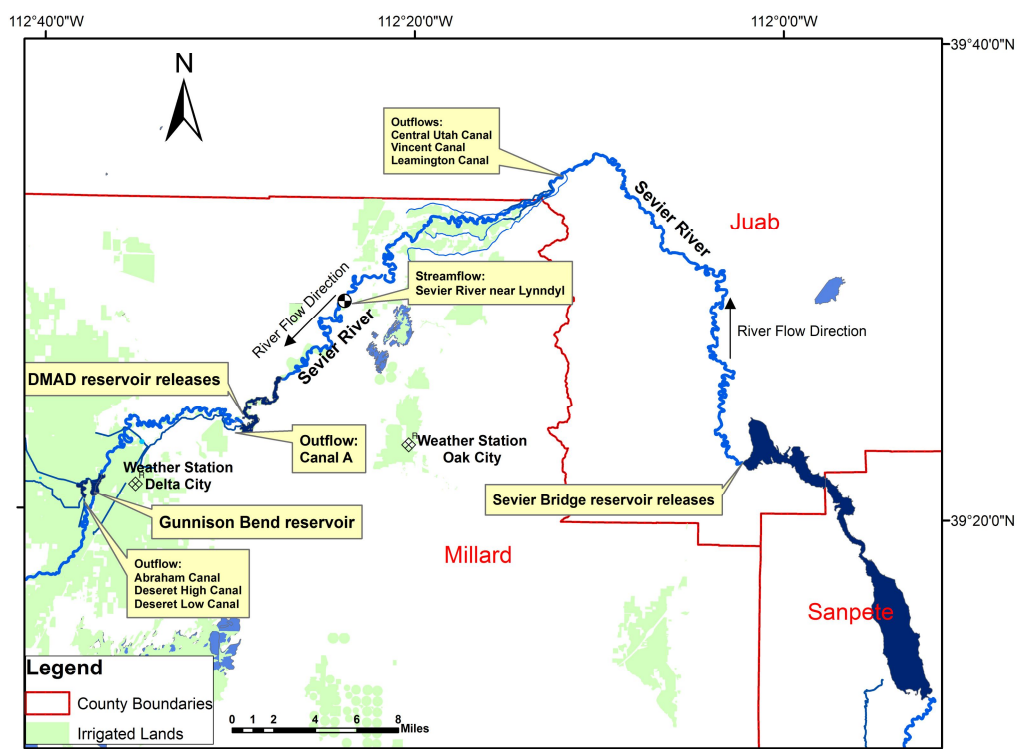
**Table 2.3** Selected ANN model for each type of training function

Testing phase								
Type of training function	Size		Statistics	R1 <sub>d</sub>	R1 <sub>d+1</sub>	R2 <sub>d</sub>	R2 <sub>d+1</sub>	Average
	of layer	Number of days <sup>1</sup>						
Quasi-Newton	3	2	E	0.942	0.867	0.898	0.777	0.871
			R	0.971	0.932	0.952	0.891	0.936
Conjugate gradient with	3	4	E	0.928	0.846	0.922	0.821	0.879
Powell-Beale restarts			R	0.964	0.922	0.961	0.908	0.939
Levenberg-Marquardt	3	5	E	0.917	0.860	0.923	0.811	0.878
			R	0.959	0.929	0.963	0.908	0.940
Scaled conjugate gradient	4	2	E	0.943	0.856	0.928	0.834	0.890
			R	0.972	0.929	0.966	0.917	0.946

<sup>1</sup> Number of prior days of historical data required as input

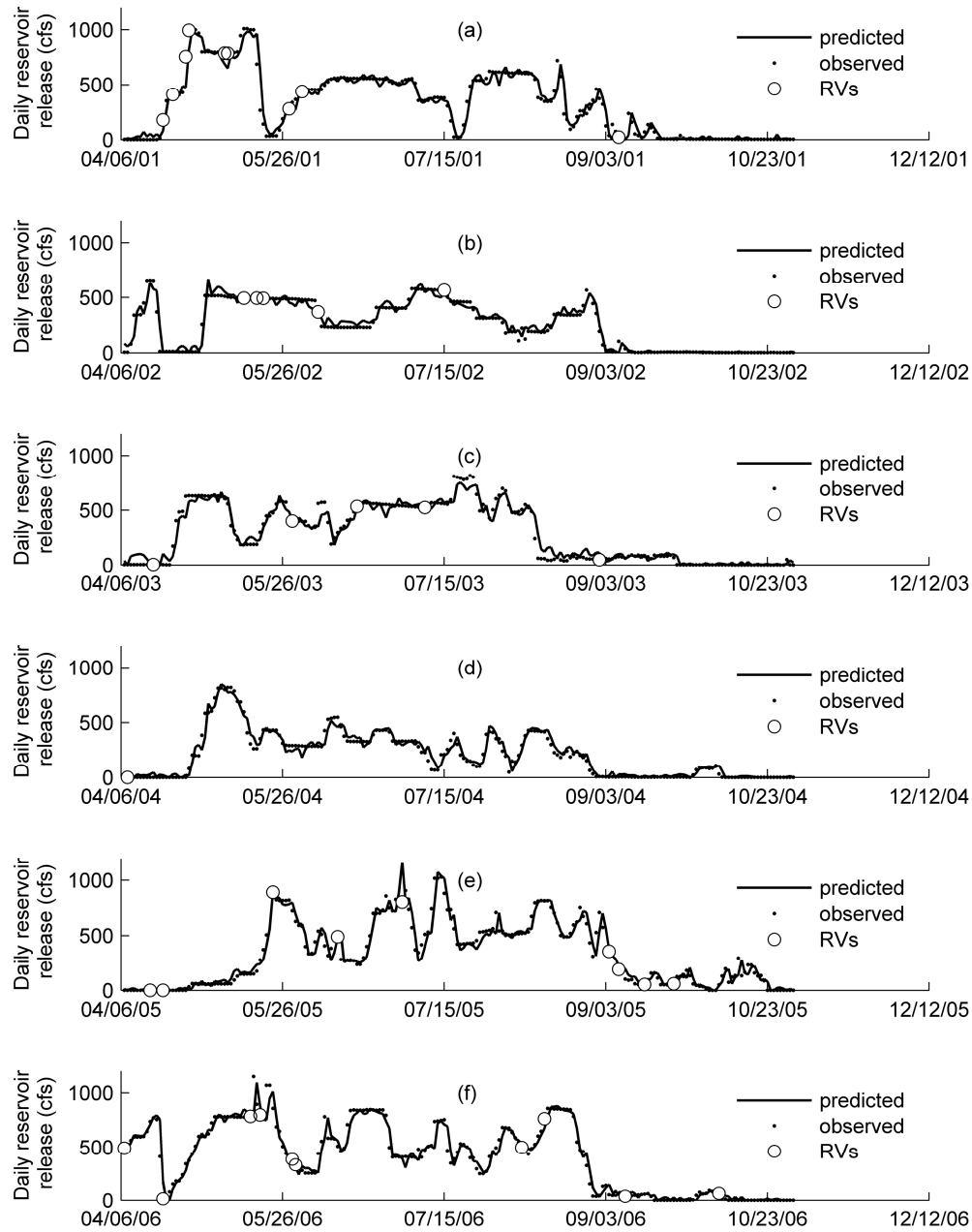
**Table 2.4** ANN performance using different statistics

Statistics	Artificial Neural Network							
	Training				Testing			
	$R1_d$	$R1_{d+1}$	$R2_d$	$R2_{d+1}$	$R1_d$	$R1_{d+1}$	$R2_d$	$R2_{d+1}$
<b>Coefficient of efficiency E</b>	0.96	0.90	0.89	0.79	0.94	0.86	0.93	0.83
<b>Correlation coefficient R</b>	0.98	0.95	0.94	0.89	0.97	0.93	0.97	0.92
<b>Root mean square error</b>								
<b>RMSE, cfs</b>	55.47	88.21	34.50	48.29	61.77	98.11	33.02	50.29

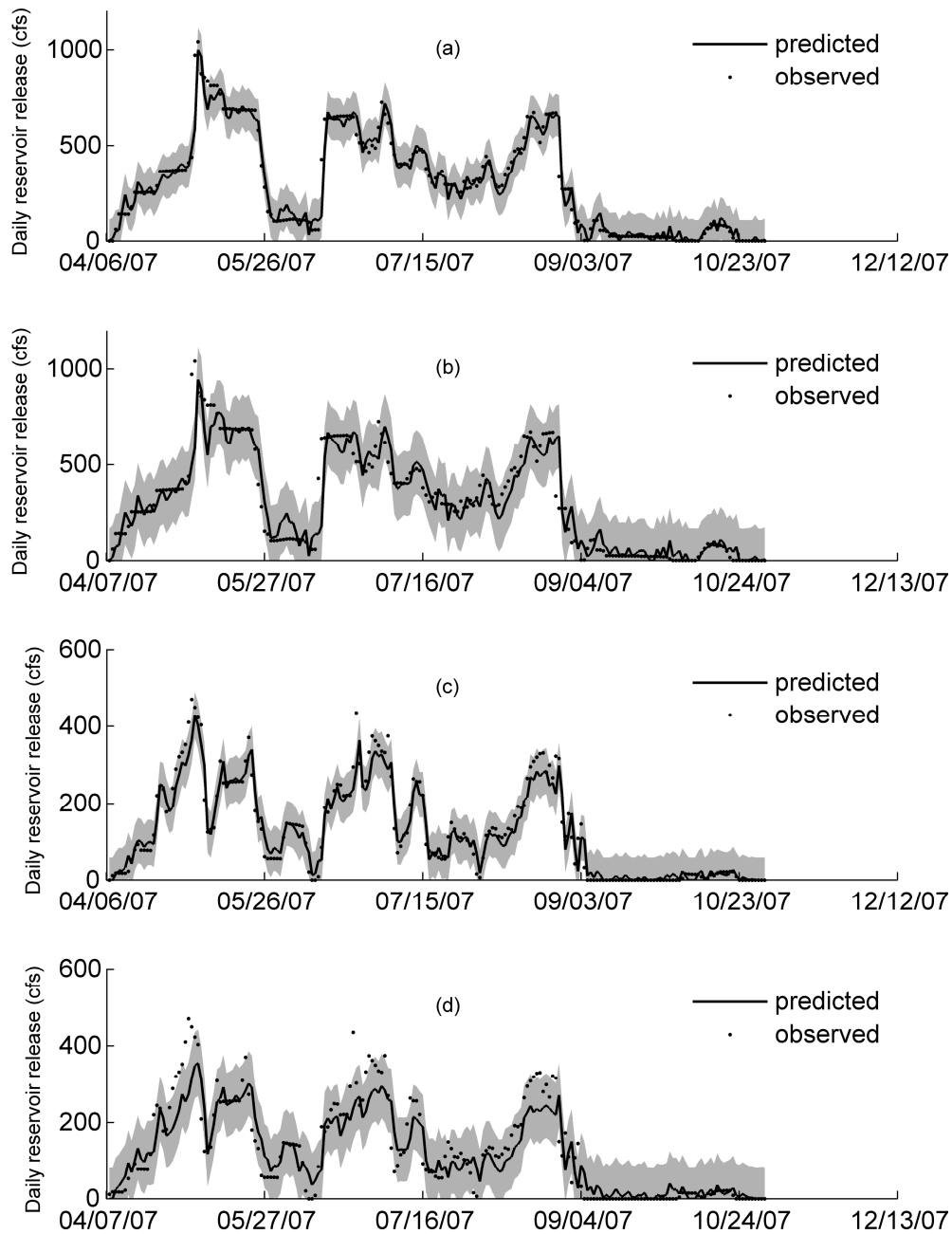


**Fig. 2.1** Lower Sevier River Basin and the spatial distribution of the sensor station locations (reservoir releases, streamflows, inflows and outflows)

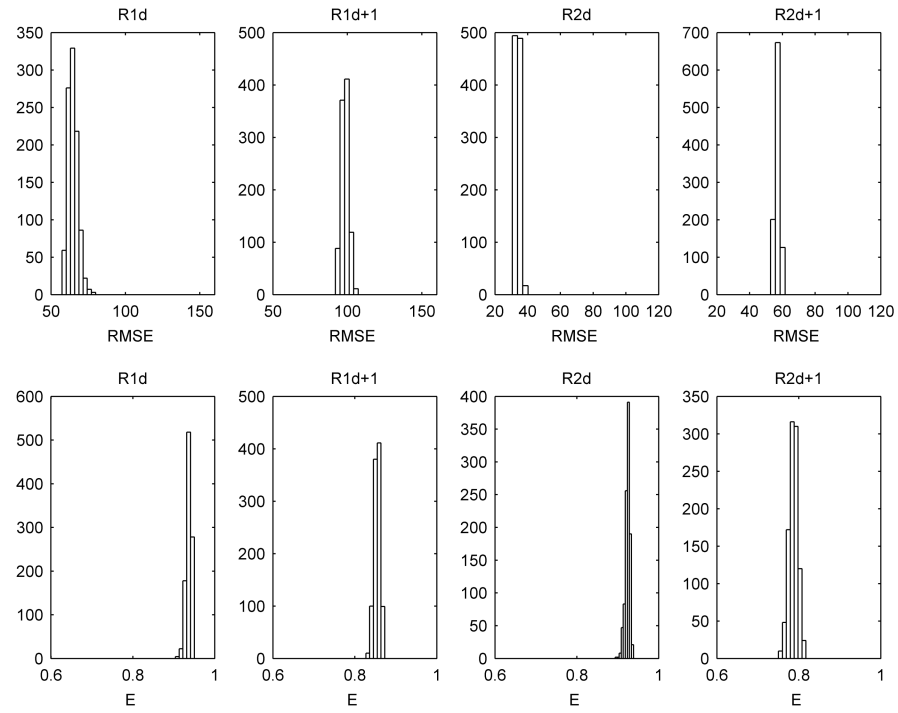




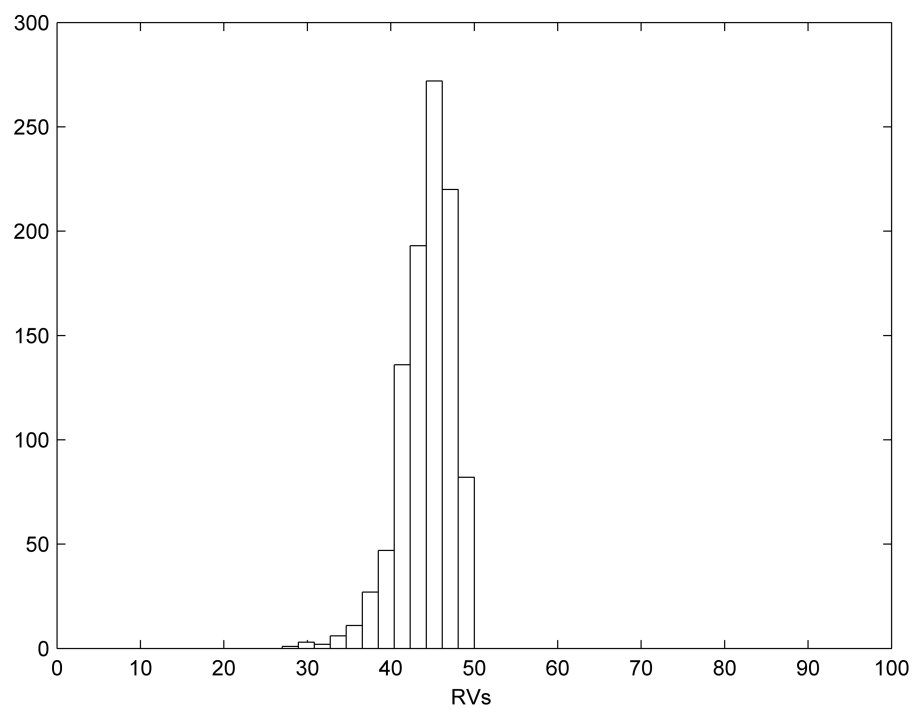
**Fig. 2.2** Plot of observed versus predicted releases of Sevier Bridge reservoir one day ahead, and RVs of the MVRVM. Training phase (2001(a) – 2006(f) irrigation seasons)



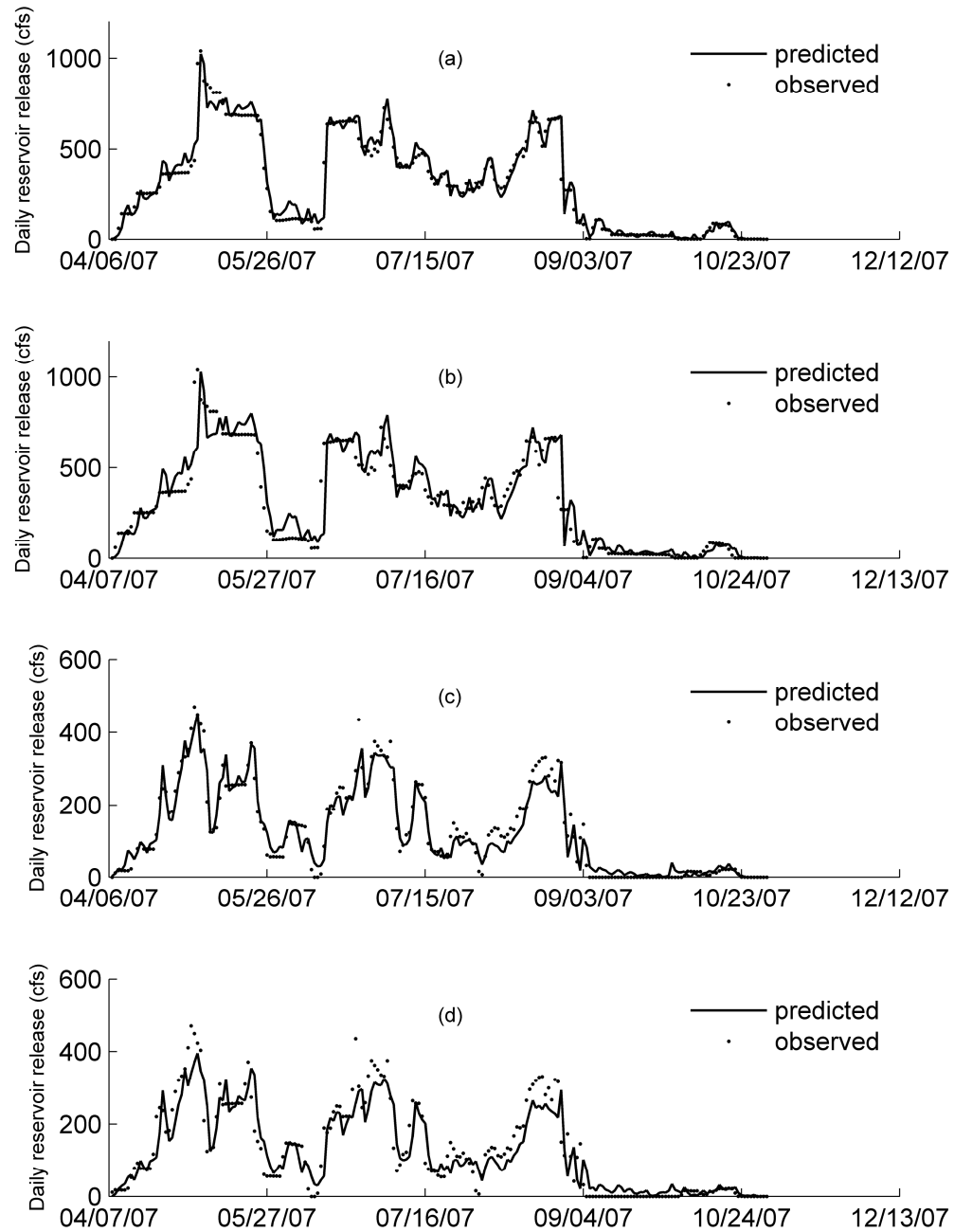
**Fig. 2.3** Observed versus predicted releases of the MVRVM with 0.90 confidence intervals (shaded region) for the testing phase (2007 irrigation season): (a) Prediction of Sevier Bridge reservoir releases one day ahead, (b) Prediction of Sevier Bridge Reservoir releases two days ahead, (c) Prediction of DMAD reservoir releases one day ahead, (d) Prediction of DMAD reservoir releases two days ahead



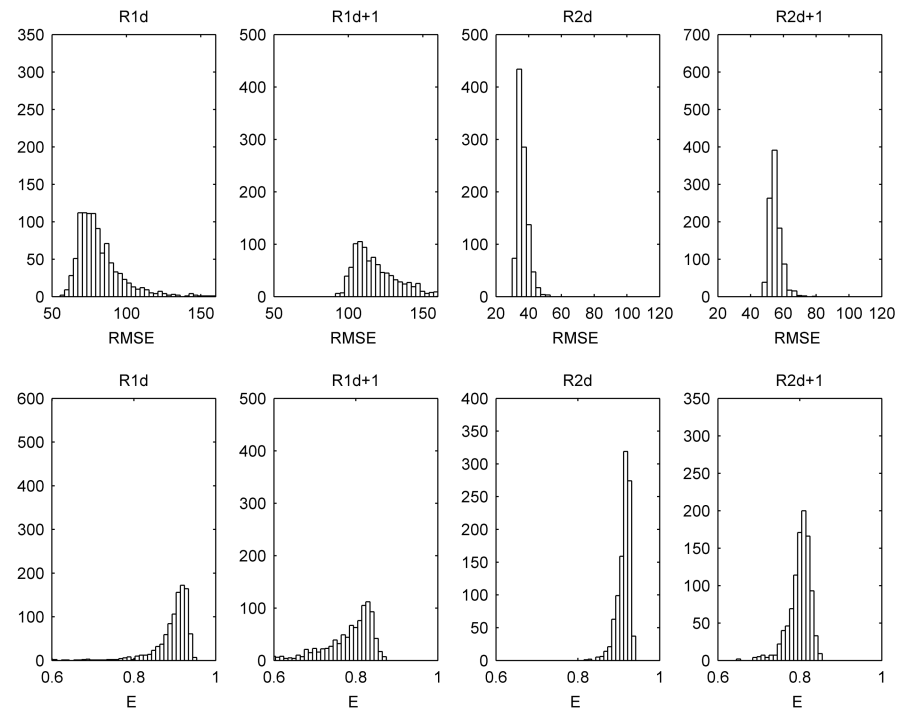
**Fig. 2.4** Bootstrap histogram of the MVRVM model for the RMSE and E test



**Fig. 2.5** Bootstrap histogram of the number of RVs



**Fig. 2.6** Observed versus predicted releases of the ANN for the testing phase (2007 irrigation season): (a) Prediction of Sevier Bridge reservoir releases one day ahead, (b) Prediction of Sevier Bridge Reservoir releases two days ahead, (c) Prediction of DMAD reservoir releases one day ahead, (d) Prediction of DMAD reservoir releases two days ahead



**Fig. 2.7** Bootstrap histogram of the ANN model for the RMSE and E test

# CHAPTER 3

## REAL-TIME FORECASTING OF SHORT-TERM IRRIGATION CANAL DEMANDS USING A ROBUST MULTIVARIATE BAYESIAN LEARNING MODEL<sup>1</sup>

### **Abstract**

This research presents models that predict the short-term diversion demands for three irrigation canals at both hourly and daily time steps. These multiple predictions will assist the operator of the reservoir located upstream of the irrigation canals, as well as the canal operators, to plan and manage in real-time the available water resources efficiently. The models are developed in the form of a multivariate relevance vector machine (MVRVM) that is based on a Bayesian learning machine approach for multivariate regression. Predictive confidence intervals can also be obtained from the model with this Bayesian approach. The models are applied to three irrigation canals located in the Lower Sevier River Basin, Utah. The inputs for the hourly model are the past hourly observations of water diversions for three canals. The outputs are the predicted diversion demands for the three canals one, twelve and twenty-four hours ahead. The inputs for the daily model are the past daily observations of water diversions for three canals and climatic data. The outputs are the predicted diversion demands for the three canals one and two days ahead. Test results show that the MVRVM learns the input-output patterns with good accuracy. A bootstrap analysis is used to evaluate robustness of model parameter estimation. The MVRVM is compared in terms of performance and robustness with an Artificial Neural Network (ANN).

---

<sup>1</sup> Coauthored by Andres M Tielavilca, Mac McKee, and Wynn Walker

### 3.1 Introduction

In recent years, the per capita availability of water resource is decreasing world-wide due to population growth, climate change, rapidly increasing demands (for irrigation, domestic supply, recreation, etc.), and pollution. Irrigated agriculture is the world's largest user of water, which accounts for 87% of consumptive uses (Pulido-Calvo and Gutierrez-Estrada 2008), and over 90% of consumptive uses in low-income developing countries (AREI 2006). This high consumption for irrigation leads to a necessity for advanced knowledge of short-term future irrigation demand so that decision makers (water operators, water managers, farmers, etc.) can achieve more efficient operation and management of water resources. The purpose of this paper is to provide a model to predict short-term future diversion demands of different irrigation canals as a function of recent historical data about diversions into the canals and climate.

Techniques based on conceptual or physical modeling have been developed to characterize the current and future states of irrigation and water resources systems. In many cases, their practical applications are limited by the lack of required data and the expense of data acquisition. This is especially critical with respect to information about model parameters which might be difficult to measure. To overcome these limitations, researchers have used data-driven modeling as an alternative to physically based models (Pulido-Calvo and Gutierrez-Estrada 2008; Khalil et al. 2005a). Examples of such models include artificial neural network (ANNs), support vector machines (SVMs), and relevance vector machines (RVMs). These types of models are derived from the emerging areas of machine learning theory. They are characterized by their ability to capture the underlying physics of the system simply by examination of the inputs and



outputs of the system. They can be used to provide predictions of the system behavior using only historical data; these models “let the data speak” (Khalil et al. 2005b).

Operation of irrigation canals is based on water requirements for crop irrigation, physical characteristic of the irrigation system, irrigation scheduling (rotational or continuous), travel times (from the diversion point to the field outlets), the agricultural area it serves (hydrologic, climatic, environmental, etc.) and human behavior, including that of the canal operator and the farmer. The canal operator has to divert water from the river into a canal to fulfill farmer’s requirements. The combination of all these behaviors and factors may cause unexpected future changes in the irrigation canal system operation. This operation could become difficult to predict. Therefore, it is necessary to develop machine learning models which have not only the ability to learn the input-output patterns and make accurate predictions but also guarantee robustness towards future changes in the system behavior.

Khalil et al. (2005b) applied a RVM model, which is based on Bayesian machine learning theory, to predict the real-time operation of a single reservoir. Their results demonstrated that the RVM model was able to predict future system states (generalization ability) and to estimate or characterize the uncertainty of the predictions (predictive confidence intervals). Another important advantage of utilizing RVMs for real-time application is their sparse formulation. RVMs typically utilize fewer basis functions when compared to SVMs (Tipping 2001). Inducing sparsity can be an effective method to control model complexity, avoid over-fitting, and control the computational characteristics of model performance (Tipping and Faul 2003).

Pulido-Calvo and Gutierrez-Estrada (2008) applied a soft-computing hybrid model for forecasting daily irrigation water demand. Their hybrid model combined feed-forward Computational Neural Networks (CNNs), fuzzy logic, and a genetic algorithm. The output of their model is the one-step-ahead prediction of the daily water demand for one irrigation canal having as inputs historical daily water demand data. Their performance results showed an accurate model.

The research reported here extends the capability introduced above to hourly and daily multi-step-ahead predictions with predictive confidence intervals for three canals simultaneously. Therefore, the models recognize the patterns between multivariate outputs (future irrigation diversion requirements for three canals) and multivariate inputs (recent data collected about the state of the system).

In order to obtain multiple-time-ahead predictions together with information that characterizes uncertainty in the prediction, the model exploits the capability of the Multivariate Relevance Vector Machine (MVRVM) (Thayananthan 2005; Thayananthan et al. 2008), a Bayesian regression tool that represents an extension of the RVM algorithm developed by Tipping and Faul (2003). It can be used to produce multivariate outputs when given a set of multivariate inputs. The MVRVM has the same capabilities of the conventional RVM: high prediction accuracy, robustness, sparse formulation, and characterization of uncertainty in the predictions. A model with all these properties can work as a practical decision support tool in real-time water resources management by providing multiple predictions that are difficult (or not practical) to obtain from traditional modeling approaches.

The remainder of this paper describes the MVRVM learning model, the area of study where the model has been applied, how the model has been developed for an irrigation canal system, the results of the MVRVM application, a comparison with an ANN model, and conclusions that can be drawn.

### 3.2 Model Description

Thayanathan (2005) proposed the Multivariate Relevance Vector Machine (MVRVM) to provide a regression tool capable of generating multivariate outputs. This model is an extension of the sparse Bayesian model developed by Tipping and Faul (2003). It is developed as follows.

Given a training data set of input-target vector pairs  $\{\mathbf{x}^{(n)}, \mathbf{t}^{(n)}\}_{n=1}^N$ , where  $N$  is the number of observations,  $\mathbf{x} \in \mathbb{R}^D$  is a  $D$ -dimensional input vector,  $\mathbf{t} \in \mathbb{R}^M$  is a  $M$ -dimensional output target vector, the model learns the dependency between input and output target with the purpose of making accurate predictions of  $\mathbf{t}$  for previously unseen values of  $\mathbf{x}$ :

$$\mathbf{t} = \Phi(\mathbf{x}) \mathbf{W} + \boldsymbol{\varepsilon} \quad (3.1)$$

where  $\mathbf{W}$  is a  $M \times P$  weight matrix and  $P = N+1$ . The error  $\boldsymbol{\varepsilon}$  is conventionally assumed to be zero-mean Gaussian with diagonal covariance matrix  $\mathbf{S} = \text{diag}(\sigma_1^2, \dots, \sigma_M^2)$ .

A fixed kernel function  $K(\mathbf{x}, \mathbf{x}_n)$  is used to create a vector of basis functions of the form  $\Phi(\mathbf{x}) = [1, K(\mathbf{x}, \mathbf{x}^{(1)}), \dots, K(\mathbf{x}, \mathbf{x}^{(N)})]$ . Tipping (2001) pointed out that this kernel function is used simply to define a set of basis functions, rather than as a definition of a dot product in some feature space such as is employed by the SVM approach. In this paper, we considered a Gaussian kernel  $K(\mathbf{x}, \mathbf{x}_n) = \exp(-r^2 \|\mathbf{x} - \mathbf{x}^{(n)}\|^2)$  where  $r$  is called the kernel width parameter, which is a smoothing parameter to define a basis function to

capture patterns in the data (Appendix A). This type of kernel has been used by several authors in water resources and hydrology applications (Khalil et al. 2005b; Tripathi and Govindaraju 2006).

Let  $\mathbf{t} = [\tau_1, \dots, \tau_r, \dots, \tau_M]^T$  and  $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_r, \dots, \mathbf{w}_M]^T$ . A likelihood distribution of the weight matrix can be written as a product of Gaussians of the weight vectors ( $\mathbf{w}_r$ ) corresponding to each target output ( $\tau_r$ ) (Thayananthan et al. 2008):

$$p(\{\mathbf{t}^{(n)}\}_{n=1}^N | \mathbf{W}, \mathbf{S}) = \prod_{n=1}^N N(\mathbf{t}^{(n)} | \mathbf{W} \Phi(\mathbf{x}^{(n)}), \mathbf{S}) = \prod_{r=1}^M N(\tau_r | \mathbf{w}_r \Phi, \sigma_r^2) \quad (3.2)$$

where  $\Phi = [1, \Phi(\mathbf{x}_1), \Phi(\mathbf{x}_2), \dots, \Phi(\mathbf{x}_N)]$ . Eq. 3.2 therefore contains the parameters  $\mathbf{W}$  (which is M by N) and  $\sigma_r^2$  (which is a vector of length M). As a result, there is a danger that the maximum likelihood estimation of  $\mathbf{w}_r$  and  $\sigma_r^2$  will suffer from severe over-fitting. To avoid this, Tipping (2001) proposed constraining the selection of parameters by applying a Bayesian perspective and defining an explicit zero-mean Gaussian prior probability distribution over them (Thayananthan et al. 2008):

$$p(\mathbf{W} | \mathbf{A}) = \prod_{r=1}^M \prod_{j=1}^P N(w_{rj} | 0, \alpha_j^{-2}) = \prod_{r=1}^M N(\mathbf{w}_r | 0, \mathbf{A}) \quad (3.3)$$

where  $\mathbf{A} = \text{diag}(\alpha_1^{-2}, \dots, \alpha_P^{-2})^T$  is a hyperparameter matrix and  $w_{rj}$  is the (r,j)th element of the weight matrix  $\mathbf{W}$ . Each  $\alpha_j$  controls the strength of the prior over its associated weight (Tipping and Faul 2003).

Bayesian inference considers the posterior distribution of the model parameters, which is proportional to the product of the likelihood and prior distributions:

$$p(\mathbf{W} | \{\mathbf{t}\}_{n=1}^N, \mathbf{S}, \mathbf{A}) \propto p(\{\mathbf{t}\}_{n=1}^N | \mathbf{W}, \mathbf{S}) p(\mathbf{W} | \mathbf{A}) \quad (3.4)$$

The posterior parameter distribution conditioned on the data can be written as the product of Gaussians for the weight vectors of each target output (Thayananthan et al. 2008):

$$p(\mathbf{W} | \{\mathbf{t}\}_{n=1}^N, \mathbf{S}, \mathbf{A}) \propto p(\{\mathbf{t}\}_{n=1}^N | \mathbf{W}, \mathbf{S}) p(\mathbf{W} | \mathbf{A}) \propto \prod_{r=1}^M \mathcal{N}(\mathbf{w}_r | \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r) \quad (3.5)$$

The posterior distribution of the weights is Gaussian  $\mathcal{N}(\mathbf{u}_r, \boldsymbol{\Sigma}_r)$  with the covariance and mean,  $\boldsymbol{\Sigma}_r = (\mathbf{A} + \sigma_r^{-2} \boldsymbol{\Phi}^T \boldsymbol{\Phi})^{-1}$  and  $\boldsymbol{\mu}_r = \sigma_r^{-2} \boldsymbol{\Sigma}_r \boldsymbol{\Phi}^T \boldsymbol{\tau}_r$ , respectively. Given this posterior, we can obtain an optimal weight matrix by getting a set of hyperparameters that maximizes the data likelihood over the weights in Eq. 3.5. The marginal likelihood is then:

$$\begin{aligned} p(\{\mathbf{t}\}_{n=1}^N | \mathbf{A}, \mathbf{S}) &= \int p(\{\mathbf{t}\}_{n=1}^N | \mathbf{W}, \mathbf{S}) p(\mathbf{W} | \mathbf{A}) d\mathbf{W}, \\ &= \prod_{r=1}^M \int \mathcal{N}(\boldsymbol{\tau}_r | \mathbf{w}_r \boldsymbol{\Phi}, \sigma_r^2) \mathcal{N}(\mathbf{w}_r | \mathbf{0}, \mathbf{A}) d\mathbf{w}_r = \prod_{r=1}^M |\mathbf{H}_r|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \boldsymbol{\tau}_r^T \mathbf{H}_r^{-1} \boldsymbol{\tau}_r\right) \end{aligned} \quad (3.6)$$

where  $\mathbf{H}_r = \sigma_r^2 \mathbf{I} + \boldsymbol{\Phi} \mathbf{A}^{-1} \boldsymbol{\Phi}^T$ . Then, we can obtain an optimal set of hyperparameters  $\boldsymbol{\alpha}^{\text{opt}} = \{\boldsymbol{\alpha}_j^{\text{opt}}\}_{j=1}^P$  and noise parameters  $(\boldsymbol{\sigma}^{\text{opt}})^2 = \{\sigma_r^{\text{opt}}\}_{r=1}^M$  by maximizing the marginal likelihood using the fast marginal likelihood maximization algorithm proposed by Tipping and Faul (2003). During the optimization process, many elements of  $\boldsymbol{\alpha}$  go to infinity, for which the corresponding posterior probability of the weight becomes zero. The relatively few nonzero weights correspond to the input vectors that form the sparse core of the RVM model. It is these input vectors, called the relevance vectors (RVs), that generate a sparse representation. Inducing sparsity can be an effective method to control model complexity, avoid over-fitting and control computational characteristics of model performance

(Tipping and Faul 2003). The optimal parameters are used to obtain the optimal weight matrix with optimal covariance  $\Sigma^{\text{opt}} = \{\Sigma_r^{\text{opt}}\}_{r=1}^M$  and mean  $\mu^{\text{opt}} = \{\mu_r^{\text{opt}}\}_{r=1}^M$ .

We can compute the predictive distribution for any new input  $\mathbf{x}^*$ , corresponding to a target  $\mathbf{t}^*$  (Tipping 2001), from:

$$p(\mathbf{t}^* | \mathbf{t}, \mathbf{a}^{\text{opt}}, (\boldsymbol{\sigma}^{\text{opt}})^2) = \int p(\mathbf{t}^* | \mathbf{W}, (\boldsymbol{\sigma}^{\text{opt}})^2) \cdot p(\mathbf{W} | \mathbf{t}, \mathbf{a}^{\text{opt}}, (\boldsymbol{\sigma}^{\text{opt}})^2) d\mathbf{W} \quad (3.7)$$

Taking into consideration that both terms in the integrand are Gaussian, Eq. 3.7 is computed as:

$$p(\mathbf{t}^* | \mathbf{t}, \mathbf{a}^{\text{opt}}, (\boldsymbol{\sigma}^{\text{opt}})^2) = \mathcal{N}(\mathbf{t}^* | \mathbf{y}^*, (\boldsymbol{\sigma}^*)^2) \quad (3.8)$$

where  $\mathbf{y}^* = [y_1^*, \dots, y_r^*, \dots, y_M^*]^T$  is the predictive mean with  $y_r^* = (\mu_r^{\text{opt}})^T \Phi(\mathbf{x}^*)$ ; and  $(\boldsymbol{\sigma}^*)^2 = [(\sigma_1^*)^2, \dots, (\sigma_r^*)^2, \dots, (\sigma_M^*)^2]^T$  is the predictive variance with  $(\sigma_r^*)^2 = (\sigma_r^{\text{opt}})^2 + \Phi(\mathbf{x}^*)^T \Sigma_r^{\text{opt}} \Phi(\mathbf{x}^*)$ .  $(\sigma_r^*)^2$  contains the sum of two variance terms: the noise on the data and the uncertainty in the prediction of the weight parameters (Tipping 2001).

The standard deviation  $\sigma_r^*$  of the predictive distribution is defined as a predictive error bar of  $y_r^*$  (Bishop 1995). Then, the width of the 90% predictive confidence interval for any  $y_r^*$  can be calculated as  $\pm 1.65 \sigma_r^*$ .

Readers interested in greater detail regarding multivariate sparse Bayesian regression, its mathematical formulation, and the optimization procedures of the model are referred to Thayanathan (2005), Thayanathan et al. (2008), Tipping (2001), and Tipping and Faul (2003). A MATLAB code developed by Thayanathan (2005) is available from <http://mi.eng.cam.ac.uk/~at315/MVRVM>.

### **3.3 Study Area**

This research deals with three irrigation canals, the Central Utah, Vincent, and Leamington Canals, that are located in the Lower Sevier River Basin near Delta, Utah (Figs. 3.1 and 3.2).

The Sevier River Basin is located in south-central Utah and is the largest drainage area in the state (approximately 12.5 percent of the state's area). Average annual precipitation ranges from 6.4 to 13.0 inches in the valleys to more than 40 inches in the high mountains. Elevation, precipitation, and temperatures are highly variable over the basin, and as a result there are several vegetative types that grow in the area. The main use of water in the basin is for irrigation. The economy of the Basin is based primarily on agriculture, tourism, and a few mining and manufacturing enterprises (Berger et al. 2003).

### **3.4 Model Application**

The Sevier River Basin is a highly instrumented and controlled basin. Automated data collection equipment (sensors on reservoirs, canals, diversions, and the river itself) was installed in the beginning of 1999 (Berger et al. 2002), and is currently collecting data at stations throughout the entire basin. The data collected are displayed in the Sevier River Water Users website, [www.sevierriver.org](http://www.sevierriver.org). This website provides the user several data retrieval and display options, such as hourly flow data for the previous 7 days or the current river and canal flow information displayed in spatial diagrams (Berger et al. 2002).

### 3.4.1 MVRVM forecasting hourly model

The irrigation canals are intensively monitored to enable real-time operation of the irrigation system. The travel time from the head gate canals to the irrigated areas ranges from 1 to 24 hours. A model that predicts hourly irrigation demands (one, twelve and twenty four hours ahead) can assist canal operators in anticipating future hourly behavior of the system and so that the system can be more intensively managed.

Hourly data from the irrigation seasons of 2005 and 2006 were used to train the MVRVM hourly model and find the model parameters. Hourly data from the 2007 irrigation season were used to test the model. The inputs are the available past hourly data collected by sensors on the canals. The multiple output target vectors are the hourly forecasts of required irrigation diversions one, 12 and 24 hours in advance for the three canals, simultaneously.

The inputs used in the model to predict hourly canal demands are expressed as

$$\mathbf{x} = [\mathbf{D1}_{h-nh}, \mathbf{D2}_{h-nh}, \mathbf{D3}_{h-nh}]^T \quad (3.9)$$

where,

$h$  = hour of prediction

$nh$  = number of hours previous to the prediction time

$\mathbf{D1}_{h-nh}$  = Central Utah Canal diversions ‘ $n$ ’ hours previous to the prediction time

$\mathbf{D2}_{h-nh}$  = Vincent Canal diversions ‘ $n$ ’ hours previous to the prediction time

$\mathbf{D3}_{h-nh}$  = Leamington Canal diversions ‘ $n$ ’ hours previous to the prediction time

The multiple output target vector of the model is expressed as

$$\mathbf{t} = [\mathbf{D1}_h, \mathbf{D1}_{h+12}, \mathbf{D1}_{h+24}, \mathbf{D2}_h, \mathbf{D2}_{h+12}, \mathbf{D2}_{h+24}, \mathbf{D3}_h, \mathbf{D3}_{h+12}, \mathbf{D3}_{h+24}]^T \quad (3.10)$$

where,



**D1<sub>h</sub>** = prediction of Central Utah Canal required diversion one hour ahead

**D1<sub>h+12</sub>** = prediction of Central Utah Canal required diversion twelve hours ahead

**D1<sub>h+24</sub>** = prediction of Central Utah Canal required diversion twenty four hours ahead

**D2<sub>h</sub>** = prediction of Vincent Canal required diversion one hour ahead

**D2<sub>h+12</sub>** = prediction of Vincent Canal required diversion twelve hours ahead

**D2<sub>h+24</sub>** = prediction of Vincent Canal required diversion twenty four hours ahead

**D3<sub>h</sub>** = prediction of Leamington Canal required diversion one hour ahead

**D3<sub>h+12</sub>** = prediction of Leamington Canal required diversion twelve hours ahead

**D3<sub>h+24</sub>** = prediction of Leamington Canal required diversion twenty four hours ahead

### 3.4.2 MVRVM forecasting daily model

The reservoir operator has to release water from Sevier Bridge reservoir in order to fulfill water demands for the irrigation canals and DMAD reservoir (Fig. 3.1). The approximate travel time between Sevier Bridge Reservoir and the canals is one day. A model that predicts required irrigation water diversions one and two days ahead can assist the operator of the Sevier Bridge Reservoir located upstream of the irrigation canals in planning and managing the available water of the reservoir efficiently.

Daily data from the 2001 through 2006 irrigation seasons were used to train the daily MVRVM model and find the model parameters. Daily data from the 2007 irrigation season were used to test the model. The inputs are the available past daily data collected by sensors on the canals and weather data.

Air temperature affects the daily rate of evapotranspiration in the irrigated areas and evaporation in the river and canals. Moreover, the daily behavior of farmers and canal operators can be directly influenced by temperature information in the basin (Khalil

et al. 2005a). Daily maximum and minimum temperature from Oak City was included in the model inputs. This historical data to build the model was obtained from The National Oceanic and Atmospheric Administration (NOAA) and recent online past data can be obtained from the National Weather Services website, [www.nws.noaa.gov](http://www.nws.noaa.gov).

The inputs used in the model to predict daily canal demands are expressed as

$$\mathbf{x} = [\mathbf{D1}_{d-nd}, \mathbf{D2}_{d-nd}, \mathbf{D3}_{d-nd}, \mathbf{Tmax}_{d-nd}, \mathbf{Tmin}_{d-nd}]^T \quad (3.11)$$

where,

$d$  = day of prediction

$nd$  = number of days previous to the prediction time

$\mathbf{D1}_{d-nd}$  = Central Utah Canal required diversion 'n' days previous to the prediction time

$\mathbf{D2}_{d-nd}$  = Vincent Canal required diversion 'n' days previous to the prediction time

$\mathbf{D3}_{d-nd}$  = Leamington Canal required diversion 'n' days previous to the prediction time

The multiple output target vector of the model is expressed as

$$\mathbf{t} = [\mathbf{D1}_d, \mathbf{D1}_{d+1}, \mathbf{D2}_d, \mathbf{D2}_{d+1}, \mathbf{D3}_d, \mathbf{D3}_{d+1}]^T \quad (3.12)$$

where,

$\mathbf{D1}_d$  = prediction of Central Utah Canal required diversion one day ahead

$\mathbf{D1}_{d+1}$  = prediction of Central Utah Canal required diversion two days ahead

$\mathbf{D2}_d$  = prediction of Vincent Canal required diversion one day ahead

$\mathbf{D2}_{d+1}$  = prediction of Vincent Canal required diversion two days ahead

$\mathbf{D3}_d$  = prediction of Leamington Canal required diversion one day ahead

$\mathbf{D3}_{d+1}$  = prediction of Leamington Canal required diversion two days ahead

### 3.4.3 Model selection

The statistic used for model selection is the coefficient of efficiency (E) calculated for the testing phase. It has been recommended by the ASCE (1993) and Legates and McCabe (1999), and is given by:

$$E = 1 - \frac{\sum_{n=1}^N (t_n - t_n^*)^2}{\sum_{n=1}^N (t_n - t_{av})^2} \quad (3.13)$$

where  $t$  is the observed output;  $t^*$  is the predicted output ;  $t_{av}$  is the observed average output and  $N$  is the number of observations. This statistic ranges from minus infinity (poor model) to 1.0 (a perfect model) (Legates and McCabe 1999).

The kernel width parameter had to be selected in order to find the appropriate model (Tipping 2001; Ghosh and Mujumdar 2007). Several MVRVM models were built with variation in kernel width and the number of previous time steps. For the daily model, the number of time steps previous to the prediction time ranges from 1 to 5 days. For the hourly model the number of time steps previous to the prediction time ranges from 4 to 48 hours. For each previous time step the kernel width ranges from 1 to 40 for both models (hourly and daily). The kernel width that is selected is the one with maximum E. From the list of models with selected kernel width at different time steps, we consider that the selected model is the one with the maximum E and minimum number of previous time steps as input data.

### 3.4.4 Bootstrap analysis

It is necessary to develop models to guarantee not only high accuracy but also good generalization and robustness of parameter estimation with respect to future changes in the nature of the input data. Changes in the training data used to build a model

may give different test results. Different sets of training data may produce models with very different accuracies. The bootstrap method was used to explore the implications of the change in the nature of input data and to guarantee good generalization ability and robustness of the MVRVM (Khalil et al. 2005b).

The bootstrap data set was created by randomly sampling with replacement from the whole training data set. Because the selection is from the whole training data set, there is nearly always duplication of individual points in a bootstrap data set. In the bootstrap estimation, this selection process was independently repeated 1,000 times to yield 1,000 bootstrap training data sets, which are treated as independent sets (Duda et al. 2001). For each of the bootstrap training data sets, a model was built and evaluated over the original test data set.

Efron and Tibshirani (1998) emphasized that it is always wise to look at the bootstrap data graphically, rather than relying entirely on a single summary statistic estimator. The bootstrap method provides information on the uncertainty in the statistics estimator evaluated in the model. According to Khalil et al. (2005b) a robust model is one that shows narrow confidence bounds in the bootstrap histograms.

#### 3.4.5 Comparison between MVRVM and ANN

ANNs have been widely applied in irrigation, water resources, and hydrologic modeling (ASCE Task Committee on the Application of ANNs in Hydrology 2000a, 2000b; Khalil et al. 2005d; Pulido-Calvo and Gutierrez-Estrada 2008; Rahimi 2008; Adeloye 2009). A comparative analysis between the MVRVM models that have been developed here and ANN models is performed in terms of performance and robustness.

Readers interested in greater detail regarding ANNs and their training functions are referred to Demuth et al. (2009),

The ANN toolbox available in MATLAB is applied in this research. The Levenberg-Marquardt (LM) algorithm was used to train the ANN model. This is a training algorithm with second-order convergence to optimize the ANN parameters (Tan and Cauwenberghe 1999) and it has been recommended by several authors in irrigation, water resources, and hydrologic research (Pulido-Calvo and Gutierrez-Estrada 2008; Rahimi 2008; Kisi 2009).

Several feed-forward ANN models were built with variation in size of the hidden layer and the number of previous time steps. The number of time steps previous to the prediction time required as input for the daily model ranges from 1 to 5 days; and the number of time steps previous to the prediction time required as input for the hourly model ranges from 4 to 48 hours. For each previous time step used as inputs, the size of the hidden layer ranges from 1 to 10. The selected hidden layer size is the one with maximum E. Similar to the MVRVM model selection and having the list of selected models at different time steps, the best model is the one with the maximum E and minimum number of previous time steps.

### **3.5 Results and Discussion**

The predicted outputs of the MVRVM for the testing phase (2007 irrigation season) are shown as the solid lines in Fig. 3.3 for the hourly model and Fig. 3.4 for the daily model. The models explain well the observed irrigation demands (dots). The performance accuracy is reduced for the three irrigation demands twenty four hours ahead and two days ahead (Figs. 3.3c and 3.4b). This accuracy reduction is found in

most of the multiple-time-ahead prediction models, where the farther we predict into the future, the less accurate the prediction becomes.

Figures 3.3 and 3.4 also show the 0.90 confidence interval (shaded region) associated with the predictive variance of the MVRVM in Eq.3.8. The confidence intervals for the twenty-hour-ahead and the two-day-ahead prediction (Figs. 3.3c and 3.4b) become wider. We can see how the uncertainty in the predictions increases when predicting further into the future for both hourly and daily models.

In the RVM approach, the relevance vectors (RVs) are subsets of the training data set that are used for prediction (Khalil et al. 2005b). As a consequence, the complexity of the model is proportional to the number of RVs. The hourly model only utilizes 26 RVs from the full training data set (3000 observations from the 2005 through 2006 irrigation seasons). The daily model utilizes 26 RVs from the full training data set (1194 observations from the 2001 through 2006 irrigation seasons). This low number of relevance vectors illustrates that the Bayesian learning procedure embodied in the MVRVM is capable of producing very sparse models.

The observed (dots) and predicted (solid lines) outputs of the ANN for the testing phase (2007 irrigation season) are shown in Fig. 3.5 for the hourly model and in Fig. 3.6 for the daily model. The models explain well the observed irrigation diversions (dots). As with the MVRVM models, the performance accuracy is reduced for the three irrigation demands two days ahead and twenty four days ahead (Figs. 3.5c and 3.6b).

Tables 3.1 and 3.2 show some statistics that measure MVRVM and ANN performances for both hourly and daily models. Again, we can see good performance of both models in the testing phase for the one-hour-ahead, twelve-hour-ahead, and one-day

ahead forecasts. The performance accuracy is reduced for both models for twenty-four-hour-ahead and two day-ahead predictions. It is important to mention that the selected models for both MVRVM and ANN required the same number of previous time steps as inputs: two days as input data for the daily models, and four hours as input data for the hourly models. We can see that the performance results for MRVM and ANN are quite similar.

Figures 3.7 and 3.8 show the bootstrap histograms of the number of RVs used in the hourly and daily MVRVM models respectively. The width of the bootstrapping confidence intervals provides information on the uncertainty in the model parameters. A narrow confidence interval (such as those illustrated in Figs. 3.7 and 3.8) implies low variability of the statistics with respect to possible future changes in the input data, which indicates that the model is robust (Khalil et al; 2005c). The low variability of the number of RVs indicates that the model structure is stable and robust with respect to future changes in the nature of the input data.

Figures 3.9, 3.10, 3.11 and 3.12 show the bootstrap histograms of the MVRVM and ANN models for the RMSE test. Most of the bootstrap histograms of the MVRVM model (Figs. 3.9 and 3.10) show very narrow confidence bounds in comparison to the histograms of the ANN model (Figs. 3.11 and 3.12). Therefore, the MVRVM is more robust.

### **3.6 Summary and Conclusions**

This paper presents a first attempt to use MVRVM models to develop multiple-time-ahead predictions of required hourly and daily diversions for a system of multiple irrigation canals. The MVRVM is a regression tool extension of the RVM algorithm to

produce multivariate outputs when given a set of multivariate inputs. The model is illustrated by application to three irrigation canals in the Lower Sevier River near Delta, Utah.

The results show that the model learns the input-output patterns with high accuracy consistent with the statistics for the test results. The performance decreases for the two-day-ahead prediction and the twenty-four-hour-ahead prediction.

The MVRVM model has the property of sparse formulation. The hourly model utilizes 26 RVs from the full data set (out of a possible 3000 observations) that was used for training. Also, the daily model only utilizes 26 RVs from the full data set that was used for training (1194 observations). The parsimonious structure of these empirical models is sufficient to explain the data and to avoid data over-fitting. Therefore, we can see an important advantage of the Bayesian learning procedure, which is the capability of the MVRVM to produce very sparse models.

An important advantage of utilizing MVRVM is its generalization capabilities. Generalization ability is associated with the capability of the model to predict future system states when presented with a range of input vectors. Irrigation canal system operation could become a difficult process to predict since this involves many decisions for real-time water resources management. The model presented here ensures good generalization providing robustness with respect to new input data.

The performance results for both the MVRVM and ANN models are fairly similar. However, most of the bootstrapped histograms of the MVRVM model show narrower confidence bounds compared to the corresponding histograms of the ANN model. Therefore, the MVRVM is more robust.



The results presented in this paper have demonstrated the successful performance and robustness of MVRVM for multiple irrigation demand forecasts. Simultaneous multiple-time-ahead demand predictions from an irrigation canal system have enormous potential value to assist the reservoir operator upstream of the canals and the canal operator in efficiently selecting the real-time operation and management decisions for available water resources.

## References

- Adeloye AJ (2009) Multiple linear regression and artificial neural networks models for generalized reservoir storage-yield-reliability function for reservoir planning. *J Hydrol Eng* 14(7):731-738.
- AREI Agricultural Resources and Environmental Indicators, Global Resources and Productivity (2006), Economic Information Bulletin No EIB-16, Washington DC.
- ASCE Task Committee on Definition of Criteria for Evaluation of Watershed Models of the Watershed Management, Irrigation, and Drainage Division (ASCE) (1993) Criteria for evaluation of watershed models. *J Irr Drain Eng* 119(3):429-442.
- ASCE Task Committee on the Application of ANNs in Hydrology (2000a) Artificial neural networks in hydrology, I: Preliminary concepts. *J Hydrol Eng* 5(2):115-123.
- ASCE Task Committee on the Application of ANNs in Hydrology (2000b) Artificial neural networks in hydrology, II: Hydrologic application. *J Hydrol Eng* 5(2):124-137.
- Berger B, Hansen R, Jensen R (2003) Sevier River Basin system description. Sevier River Water Users Association, Delta, Utah.
- Bishop CM (1995) Neural network for pattern recognition. Oxford University Press, USA
- Demuth H, Beale M, Hagan M (2009) Neural Network Toolbox user's guide. The MathWorks Inc, USA.
- Duda RO, Hart P, Stork D (2001) Pattern classification. Second Edition. Wiley Interscience, USA.

- Efron B, Tibshirani R (1998) An introduction of the Bootstrap. Monographs on Statistics and Applied Probability 57, CRC Press LLC, USA.
- Ghosh S, Mujumdar P P (2007) Statistical downscaling of GCM simulations to streamflow using relevance vector machine. *Adv in Water Resours* 31:132-146.
- Khalil A, Almasari M, McKee M, Kemblowski MW, Kaluarachchi J (2005a) Applicability of statistical learning algorithms in groundwater quality modeling. *Water Resour Res* 41:W05010.
- Khalil A, McKee M, Kemblowski MW, Asefa T (2005b) Sparse Bayesian learning machine for real-time management of reservoir releases. *Water Resour Res* 41:W11401.
- Khalil A, McKee M, Kemblowski MW, Asefa T, Bastidas L (2005c) Multiobjective analysis of chaotic dynamic systems with sparse learning machines. *Advances in Water Resources* 29: 72-88.
- Khalil A, McKee M, Kemblowski MW, Asefa T (2005d) Basin-scale water management and forecasting using neural networks. *J Am Water Resour Res* 41:195-208.
- Kisi O (2009) Daily pan evaporation modeling using multi-layer perceptrons and radial basis neural networks. *Hydrol Process* 23:213-223.
- Legates D R, and McCabe G J (1999) Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. *Water Resour Res* 35(1):233-241.
- Pulido-Calvo I, Gutierrez-Estrada J C (2008) Improved irrigation water demand forecasting using a soft-computing hybrid model. *Biosyst Eng* 102(2):202-218.
- Rahimi A K (2008) Artificial neural network estimation of reference evapotranspiration from pan evaporation in a semi-arid environment. *Irrig Sci* 27:35-39.
- Tan T, Van Cauwenberghe A (1999) Neural-network-based d-step-ahead predictor for nonlinear systems with time delay. *Eng Appl of Artif Intell* 12:21-35.
- Thayananthan A (2005) Template-based Pose Estimation and Tracking of 3D Hand Motion. Dissertation, University of Cambridge, UK.
- Thayananthan A, Navaratnam R, Stenger B, Torr P H S, Cipolla R (2008) Pose estimation and tracking using multivariate regression. *Pattern Recognit Lett* 29(9):1302-1310.
- Tipping M E (2001) Sparse Bayesian learning and the relevance vector machine. *J Mach Learn* 1:211–244.

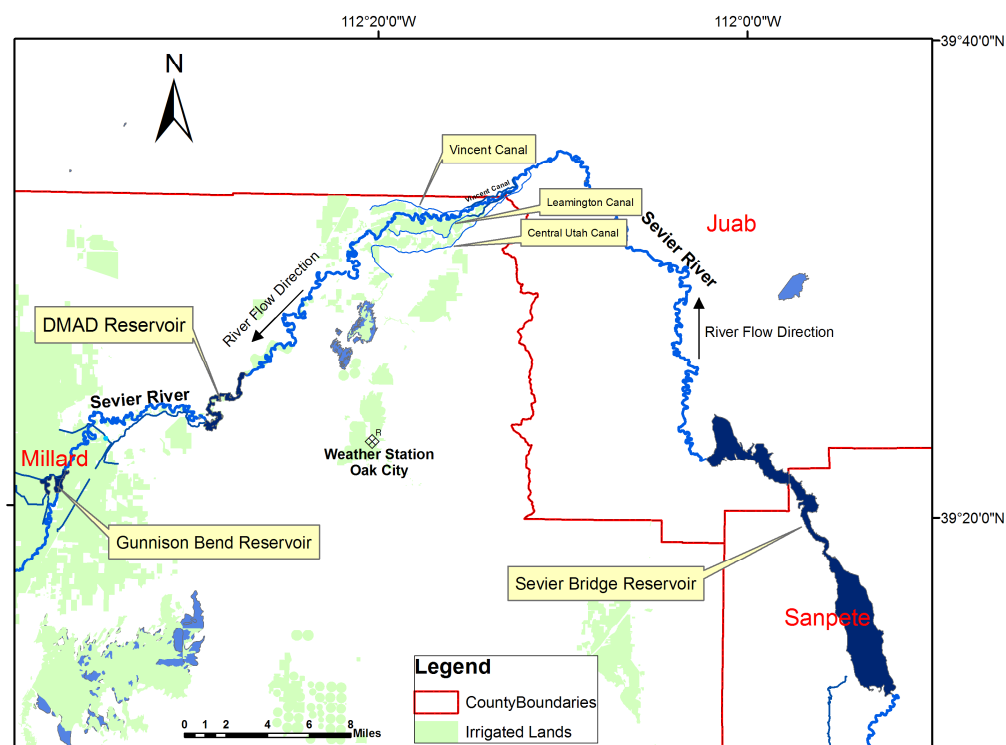
- Tipping M, Faul A (2003) Fast marginal likelihood maximization for sparse Bayesian models. Paper presented at Ninth International Workshop on Artificial Intelligence and Statistics, Soc. for Artif. Intel. and Stat., Key West, Florida.
- Tripathi S, Govindaraju R (2006) On selection of kernel parameters in relevance vector machines for hydrologic applications. *Stoch Environ Res Risk Assess* 21:747-764.

**Table 3.1** Hourly model performance using different statistics for the testing phase

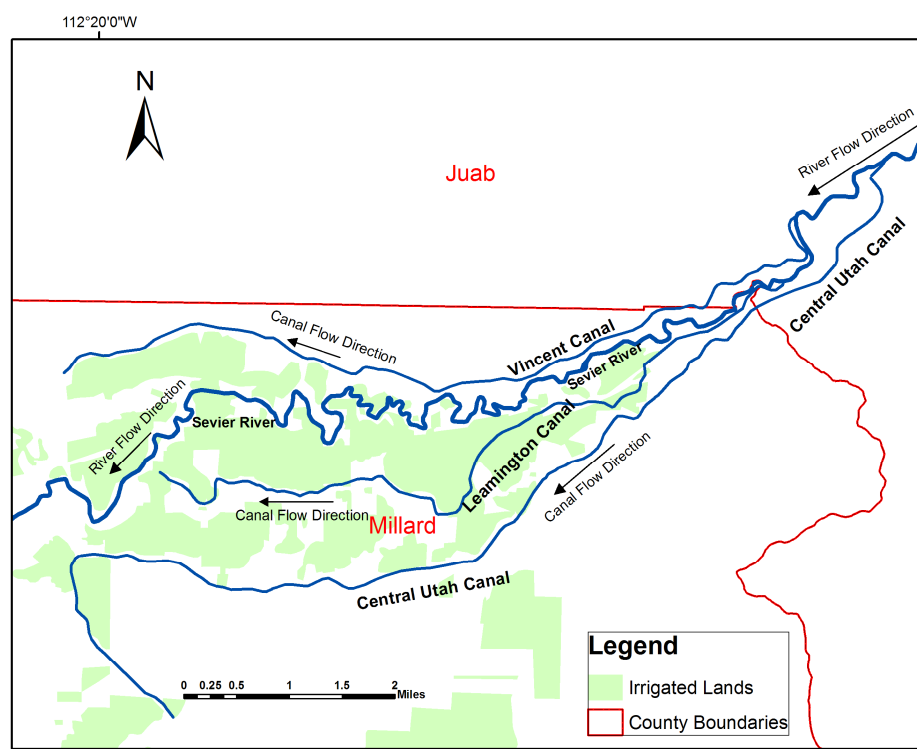
Model	Statistics	Central Utah Canal			Vincent Canal			Leamington Canal		
		1-hour	12-hours	24-hours	1-hour	12-hours	24-hours	1-hour	12-hours	24-hours
<b>MVRVM</b> (selected kernel width = 14)	<b>E</b>	0.993	0.898	0.785	0.996	0.942	0.881	0.986	0.848	0.736
	<b>RMSE, cfs</b>	2.879	10.913	15.824	0.549	2.177	3.136	1.460	4.864	6.415
	<b>Predictive error, cfs</b>	4.165	13.334	18.385	0.790	1.963	2.582	1.549	5.126	6.810
<b>ANN</b> (selected hidden layer = 5)	<b>E</b>	0.994	0.899	0.779	0.996	0.943	0.881	0.989	0.853	0.748
	<b>RMSE, cfs</b>	2.682	10.821	16.040	0.562	2.171	3.131	1.336	4.780	6.266

**Table 3.2** Daily model performance using different statistics for the testing phase

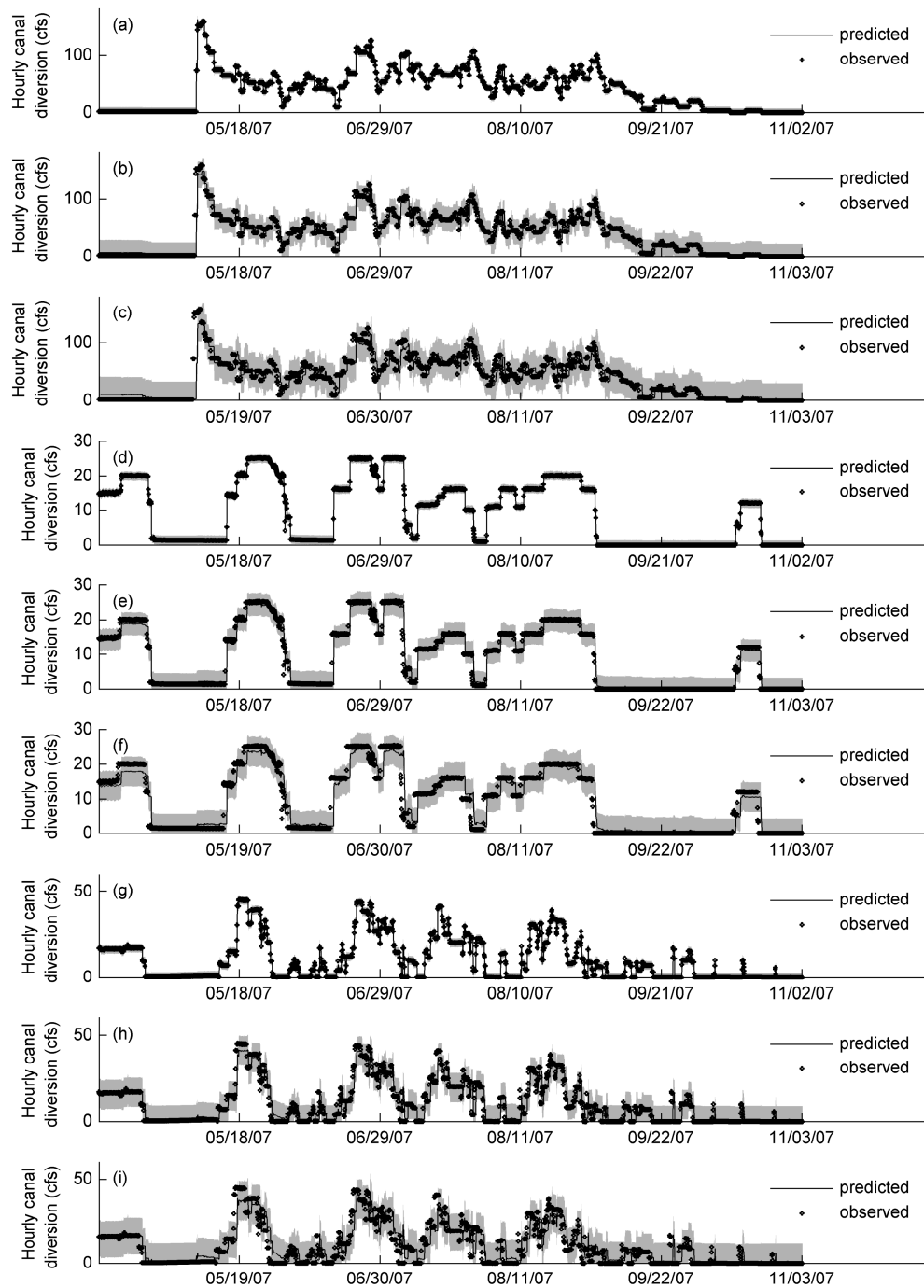
Model	Statistics	Central Utah Canal		Vincent Canal		Leamintong Canal	
		1-day	2-days	1-day	2-days	1-day	2-days
MVRVM (selected kernel width = 25)	E	0.851	0.659	0.927	0.803	0.846	0.671
	RMSE, cfs	12.998	19.648	2.515	4.116	4.890	7.158
	Predictive error, cfs	15.676	23.270	3.227	4.700	4.912	7.204
ANN (selected hidden layer size = 5)	E	0.840	0.658	0.927	0.809	0.831	0.659
	RMSE, cfs	13.447	19.684	2.514	4.056	5.122	7.293



**Fig. 3.1** Lower Sevier River Basin

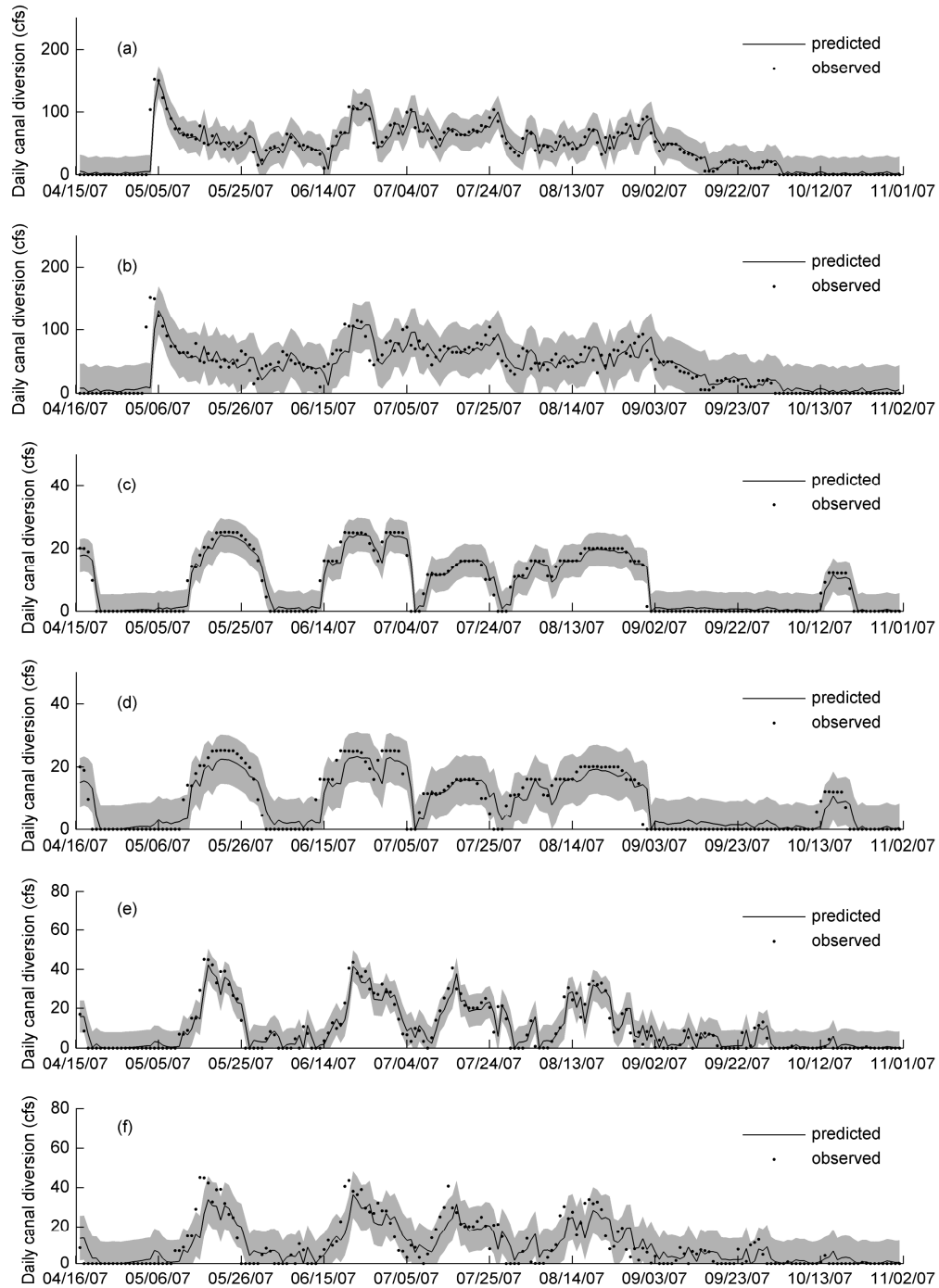


**Fig. 3.2** Sevier River, Central Utah Canal, Vincent Canal, and Leamington Canal

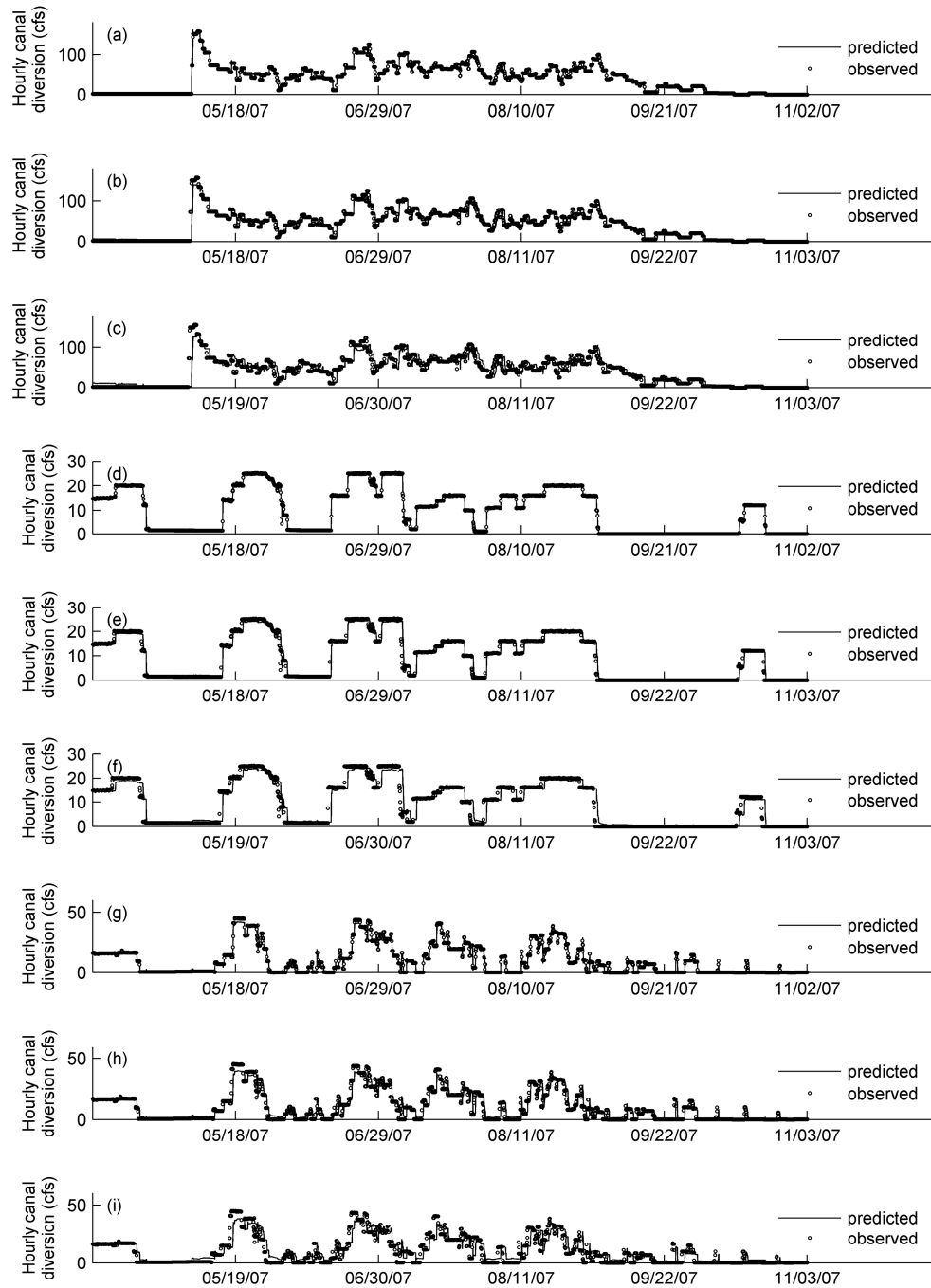


**Fig. 3.3** Observed versus predicted diversions for the hourly MVRVM model with 0.90 confidence intervals (shaded region) for the testing phase (2007 irrigation season). Central Utah Canal: (a) 1-hour ahead, (b) 12-hours ahead, (c) 24-hours ahead; Vincent Canal: (d) 1-hour ahead, (e) 12-hours ahead, (f) 24-hours ahead; Leamington Canal: (g) 1-hour ahead, (h) 12-hours ahead, (i) 24-hours ahead

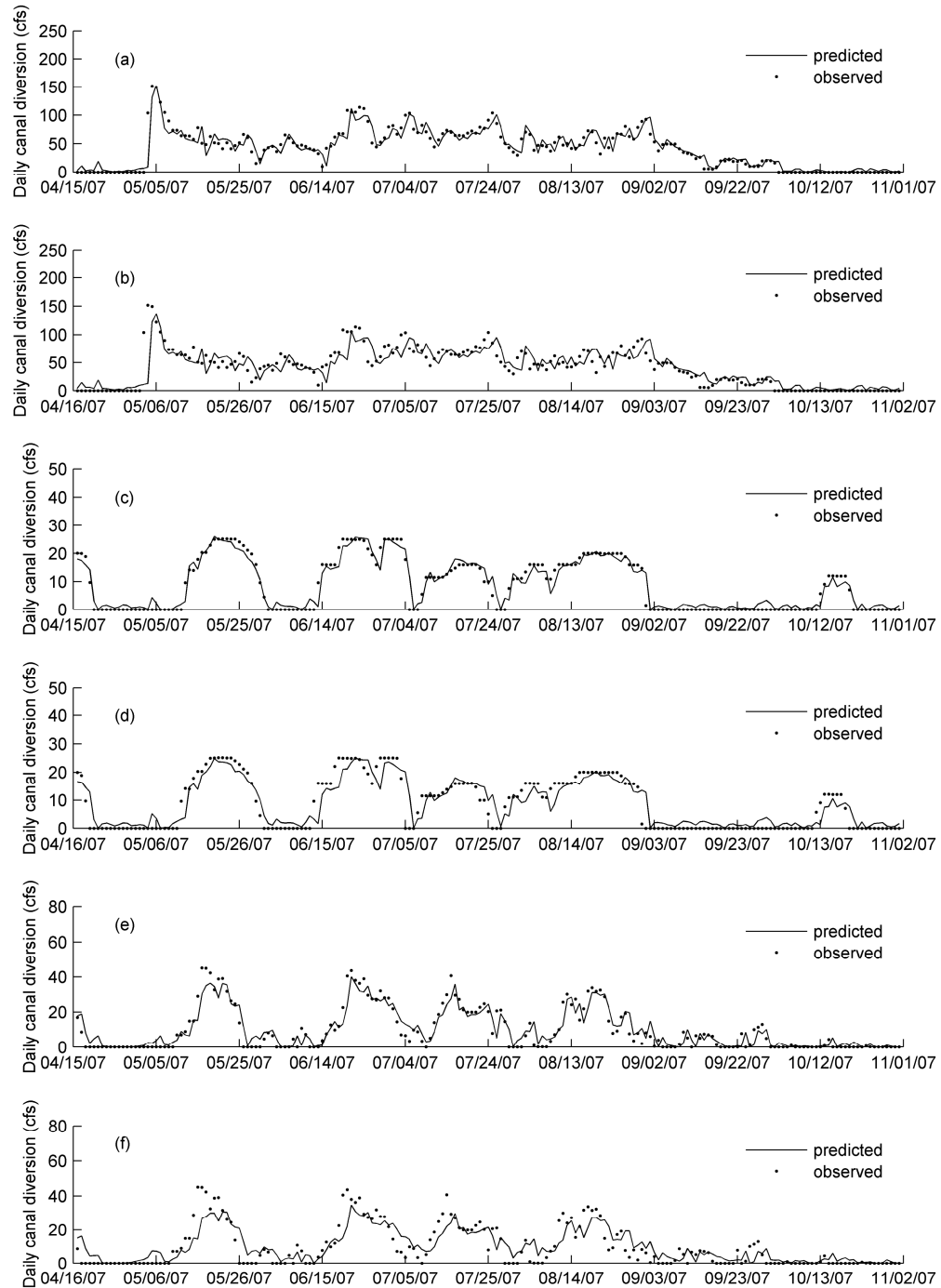




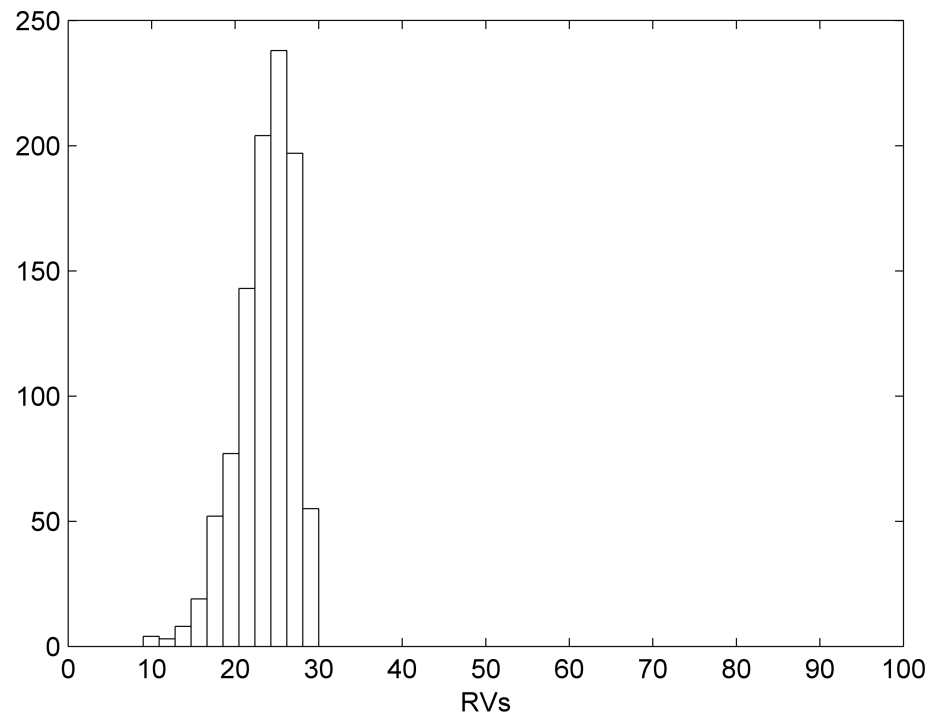
**Fig. 3.4** Observed versus predicted diversions of the daily MVRVM model with 0.90 confidence intervals (shaded region) for the testing phase (2007 irrigation season). Central Utah Canal: (a) 1-day ahead, (b) 2-days ahead; Vincent Canal: (c) 1-day ahead, (d) 2-days ahead; Leamington Canal: (e) 1-day ahead, (f) 2-days ahead



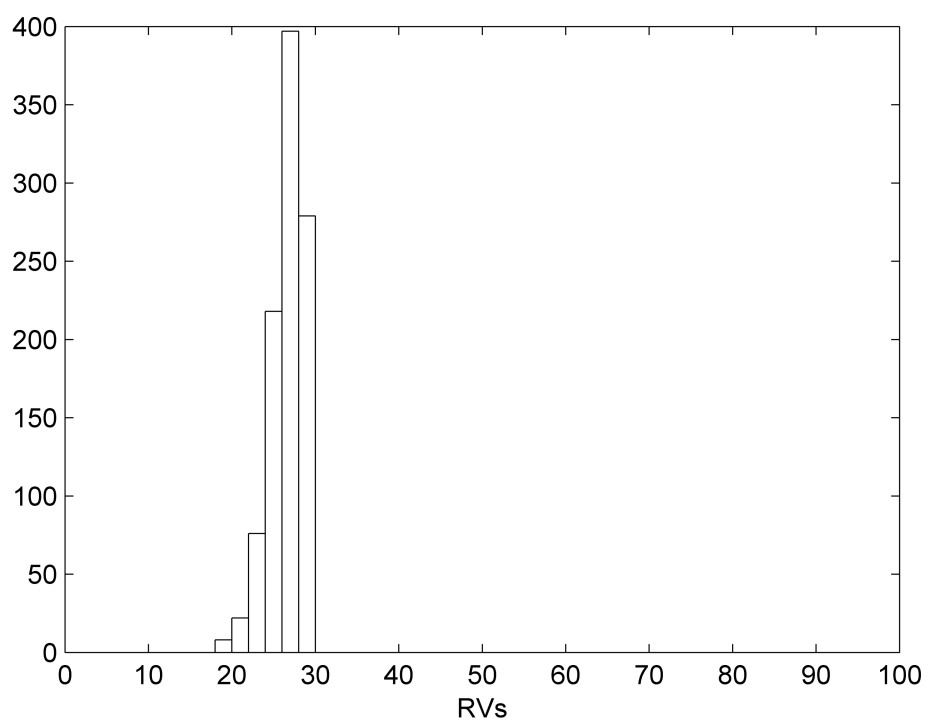
**Fig. 3.5** Observed versus predicted diversions of the hourly ANN model for the testing phase (2007 irrigation season). Central Utah Canal: (a) 1-hour ahead, (b) 12-hours ahead, (c) 24-hours ahead; Vincent Canal: (d) 1-hour ahead, (e) 12-hours ahead, (f) 24-hours ahead; Leamington Canal: (g) 1-hour ahead, (h) 12-hours ahead, (i) 24-hours ahead



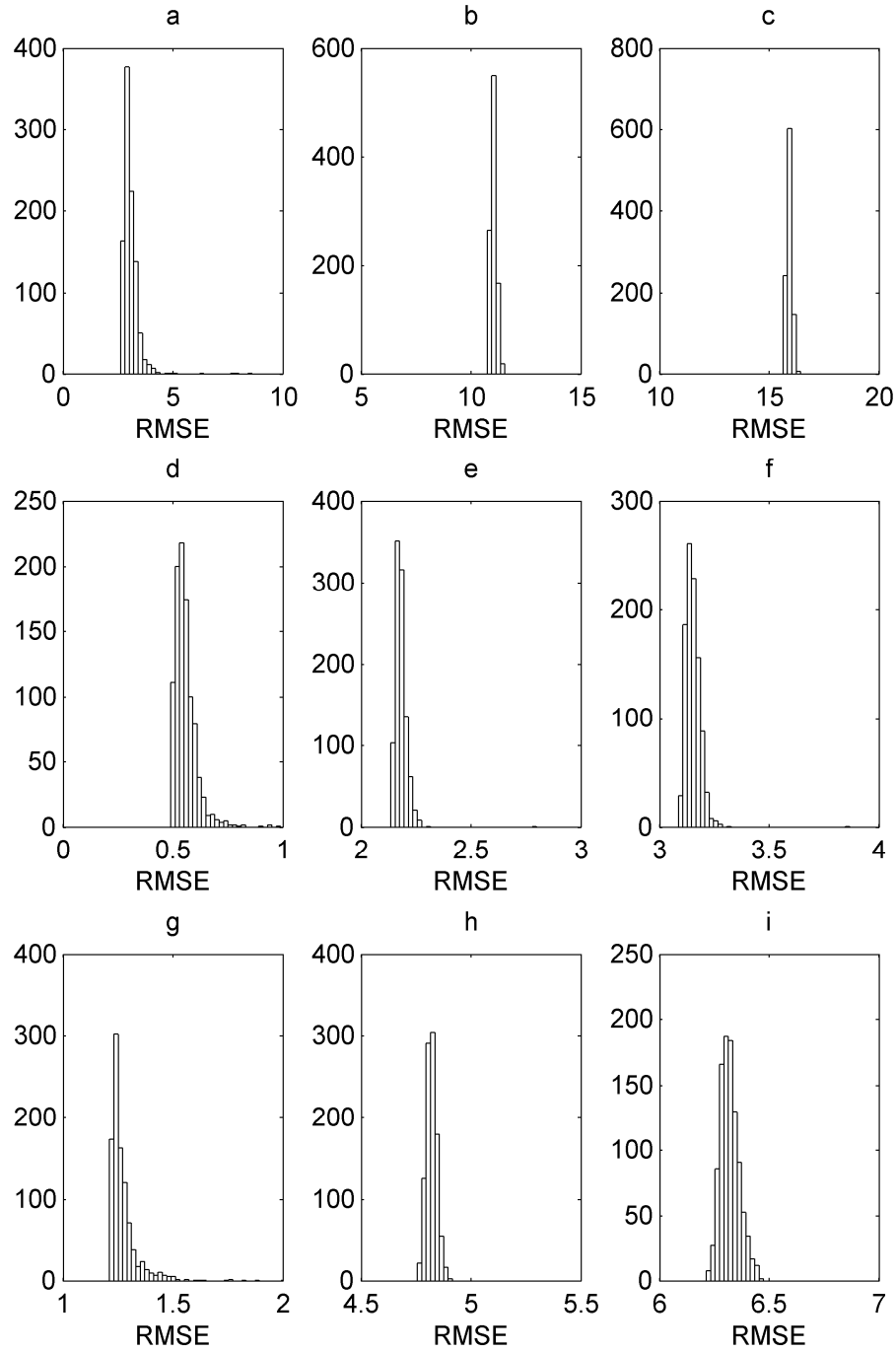
**Fig. 3.6** Observed versus predicted diversions of the daily ANN model for the testing phase (2007 irrigation season). Central Utah Canal: (a) 1-day ahead, (b) 2-days ahead; Vincent Canal: (c) 1-day ahead, (d) 2-days ahead; Leamington Canal: (e) 1-day ahead, (f) 2-days ahead



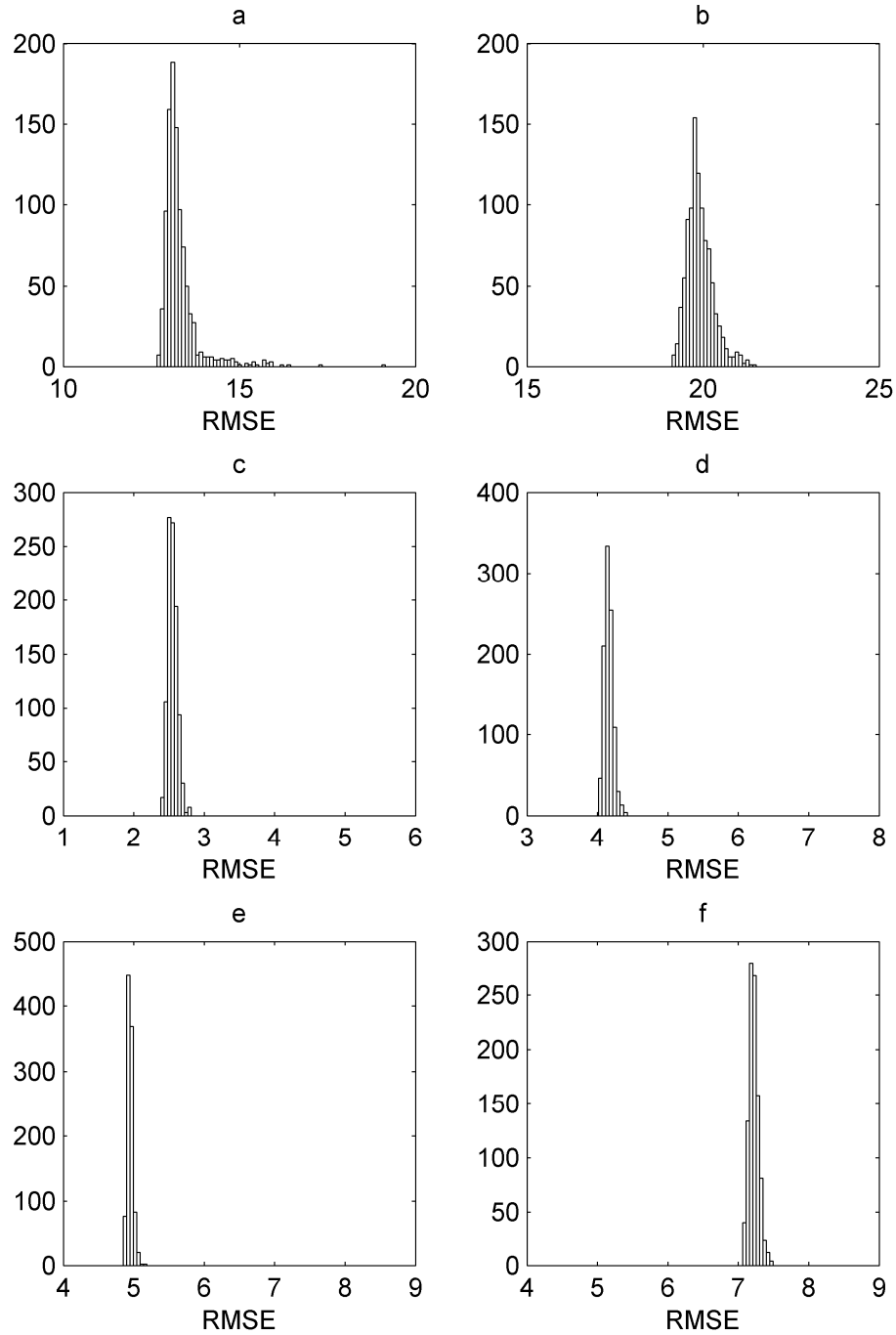
**Fig. 3.7** Bootstrap histogram of the hourly MVRVM model for the number of RVs



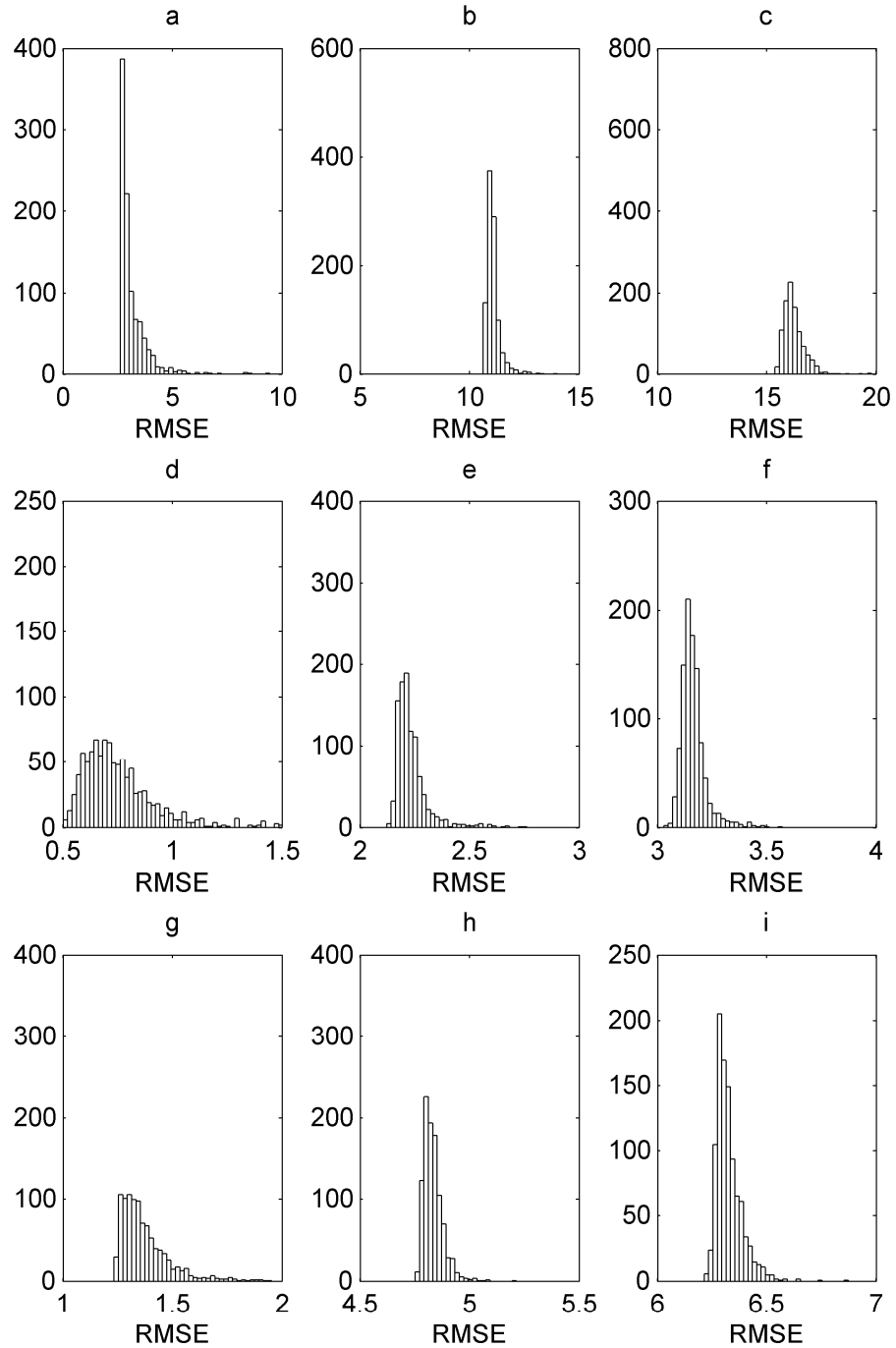
**Fig. 3.8** Bootstrap histogram of the daily MVRVM model for the number of RVs



**Fig. 3.9** Bootstrap histograms of the hourly MVRVM model for the RMSE test. Central Utah Canal: (a) 1-hour ahead, (b) 12-hours ahead, (c) 24-hours ahead; Vincent Canal: (d) 1-hour ahead, (e) 12-hours ahead, (f) 24-hours ahead; Leamington Canal: (g) 1-hour ahead, (h) 12-hours ahead, (i) 24-hours ahead

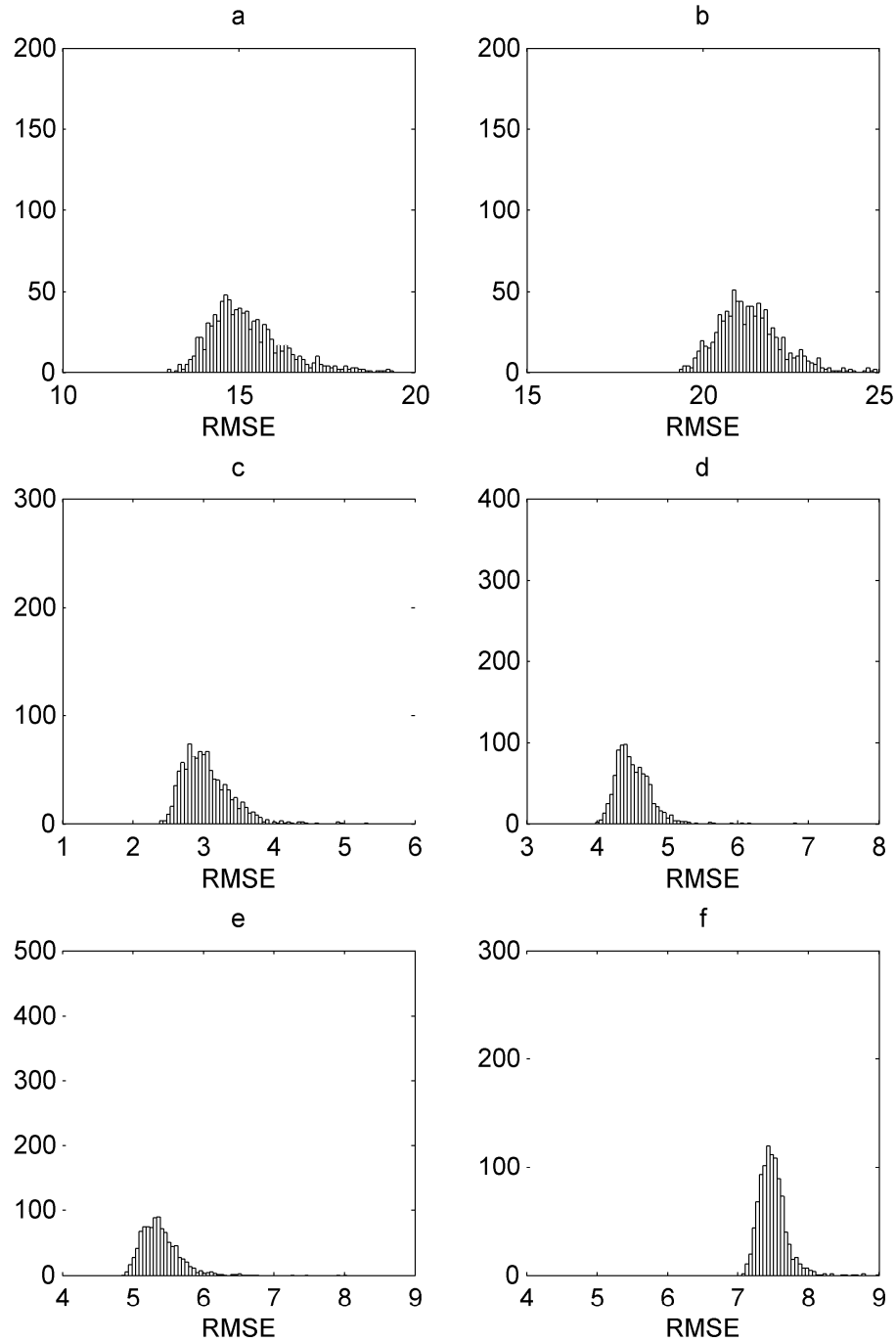


**Fig. 3.10** Bootstrap histograms of the daily MVRVM model for the RMSE test. Central Utah Canal: (a) 1-day ahead, (b) 2-days ahead; Vincent Canal: (c) 1-day ahead, (d) 2-days ahead; Leamington Canal: (e) 1-day ahead, (f) 2-days ahead



**Fig. 3.11** Bootstrap histograms of the hourly ANN model for the RMSE test. Central Utah Canal: (a) 1-hour ahead, (b) 12-hours ahead, (c) 24-hours ahead; Vincent Canal: (d) 1-hour ahead, (e) 12-hours ahead, (f) 24-hours ahead; Leamington Canal: (g) 1-hour ahead, (h) 12-hours ahead, (i) 24-hours ahead





**Fig. 3.12** Bootstrap histograms of the daily ANN model for the RMSE test. Central Utah Canal: (a) 1-day ahead, (b) 2-days ahead; Vincent Canal: (c) 1-day ahead, (d) 2-days ahead; Leamington Canal: (e) 1-day ahead, (f) 2-days ahead

CHAPTER 4

REAL-TIME FORECASTING OF STREAMFLOW AND WATER LOSS/GAIN IN  
A RIVER SYSTEM USING A MULTIVARIATE BAYESIAN REGRESSION  
MODEL<sup>1</sup>

ABSTRACT

This research presents a model that simultaneously forecasts streamflow one and two days into the future, and net cumulative water loss/gain in a river reach over the same two-day period. The model is applied to a river reach between two reservoirs located in the Lower Sevier River Basin near Delta, Utah. The reservoir operator can take into account these real-time predictions and decide how to manage releases from the upper reservoir into the lower one. The model inputs are the past daily data of climate (maximum and minimum temperature), streamflow, reservoir releases, water loss/gain in the river, and irrigation canal diversions. The model is developed in the form of a multivariate relevance vector machine (MVRVM) that is based on a multivariate Bayesian regression approach. Using this Bayesian approach, a predictive confidence interval is obtained from the model that captures the uncertainty of both the model and the data. Test results show that the MVRVM model learns the input-output patterns with good prediction accuracy. Computing multiple-time-ahead predictions in real-time for river systems would require a model which guarantees not only good prediction accuracy but also robustness with respect to future changes in the nature of the multivariate input data. A bootstrap analysis is used to evaluate robustness of model parameter estimation. The MVRVM is compared with an Artificial Neural Network (ANN) in terms of performance and robustness.

---

<sup>1</sup> Coauthored by Andres M Tielavilca, Mac McKee, and Wynn Walker

#### 4.1 INTRODUCTION

The demand for water is increasing (for irrigation, domestic supply, recreation, etc.) while the availability of water is decreasing world-wide due to population growth, climate change, and pollution. This mismatch between water demand and available usable water leads to a necessity for more intensive management of this increasingly scarce resource. That is why real-time information on the future patterns of water availability should receive particular consideration so that decision makers (water operators, water managers, farmers, etc.) can be better informed for efficient operation of water systems.

A critical problem in managing river systems is the unpredictable changes of streamflow and water loss/gain in a river reach. This is critical when water loss/gain occurs in the reach between two reservoirs (that are in series) and the reservoir operator needs to manage how much water should enter the reservoir downstream. Also, unpredictable changes in streamflow in large river basins are critical if they are related to the incidence of floods and drought in the river basin.

The main factors that influence the streamflow and water loss/gain in a river reach are return flows, transmission losses, tributary contributions, stream channel characteristics, and hydraulic characteristics of the aquifer (Reddy and Wilamowski, 2000; Jothiprakash, 2004). Conceptual and physical based models are generally used to simulate and estimate streamflow and water loss/gain. Jothiprakash (2004) developed a water balance model to estimate the water loss/gain in a river system in India. In his study, the return flow, tributary contribution, and transmission losses were assumed as a percentage of releases made from reservoirs. Hurkmans *et al.* (2008) compared the

accuracy of a conceptual water balance model (using temperature and precipitation as input data) and a more detailed physical based land surface model for streamflow simulations. Their results showed that the physical model outperforms the conceptual model during the validation period. However, this physical model required more atmospheric input data and relatively long computation times.

Developing a physical model to provide real-time streamflow and water loss/gain forecasts may not be practical if most of the data required for such a model are not available or are limited by the expense of data acquisition. To overcome these limitations, researchers have used top-down or data-driven modeling as an alternative to traditional models (Pulido-Calvo and Gutierrez-Estrada, 2008; Khalil *et al.*, 2005a; Ivkovic, 2009). Examples of such models include artificial neural networks (ANNs), support vector machines (SVMs) and relevance vector machines (RVMs). These types of models are derived from the emerging area of machine learning theory. In these modeling approaches, a model is formulated to capture the pattern between inputs and outputs of a physical system from commonly available data on inputs and outputs (Pulido-Calvo and Gutierrez-Estrada, 2008; Solomatine and Shrestha, 2009).

Several machine learning models to predict streamflow have been reported. Reddy and Wilamowski (2000) applied ANNs to forecast streamflows in order to meet downstream demands from a reservoir. Their results showed good performance in terms of matching the supply with demand. Khalil *et al.* (2005d) used ANNs to predict future streamflows at different timescales. Birikundavyi *et al.* (2002) demonstrated that ANNs outperform deterministic models for forecasting daily streamflows. Wu *et al.* (2005) applied ANNs for watershed runoff and streamflow forecasting. Ghosh and Mujumdar

(2007) demonstrated the advantage of RVMs over SVMs to model streamflow at the river basin scale for the monsoon period using simulated climate variables. Kisi (2007) performed a comparison of different ANNs for daily streamflow forecasting.

Streamflow and water loss/gain in the river reach could be difficult processes to model since they are influenced by many hydrological factors and processes (return flow, hydraulic characteristics of the aquifer, etc.). The combination of all these factors may cause unexpected future changes in the streamflow and water loss/gain in a river reach. These processes could become difficult to predict and (more importantly) to generalize. Therefore, this paper also seeks to develop machine learning models which not only have the ability to make accurate predictions but also guarantee a robust model which generalizes well towards future changes in the system behavior.

Bayesian machine learning can be used, via its probabilistic approach, to avoid overfitting during the parameter estimation process and to guaranty generalization performance. Khalil *et al.* (2005b) applied a RVM model, which is based on Bayesian machine learning theory, to predict the real-time operation of a single reservoir. Their results demonstrated that the RVM model was able to predict future system states (generalization ability) and had the capability to estimate the uncertainty of the predictions in the form of predictive confidence intervals. RVM utilizes, via its sparse formulation, fewer basis functions when compared to SVMs (Tipping, 2001). Inducing sparsity can be an effective method to control model complexity, avoid overfitting, and control the computational characteristics of a model performance (Tipping and Faul, 2003).

The purpose of this paper is to provide a robust machine learning model to forecast simultaneously streamflow one and two days ahead, and net cumulative water loss/gain in a river over one and two days. These simultaneous and multiple predictions are functions of recent climate, streamflow, reservoir releases, water loss/gain in the river, and irrigation canal diversion data. Therefore, the model recognizes the patterns between multivariate outputs (future streamflow and water loss/gain) and multivariate inputs (past observations collected from the system).

In order to obtain multiple-time-ahead predictions with predictive confidence intervals, the model exploits the capability of the Multivariate Relevance Vector Machine (MVRVM) (Thayananthan, 2005; Thayananthan *et al.*, 2008). The MVRVM is a Bayesian regression tool extension of the RVM algorithm developed by Tipping and Faul (2003) to produce multivariate outputs when given a set of inputs. In addition to its ability to predict multiple outputs, the MVRVM has the same properties of the conventional RVM: high prediction accuracy, robustness, sparse formulation, and characterization of uncertainty in the predictions. Therefore, developing a model with all these properties can work as a practical decision support tool in real-time river system management by providing multiple predictions that are difficult (or not practical) to obtain from traditional modeling approaches.

The remainder of the paper describes the MVRVM learning model, the area of study where the model has been applied, how the model has been developed for a river system, the results of the MVRVM application, a comparison with an ANN model, and conclusions that can be drawn.

## 4.2 Material and Methods

### 4.2.1 Model description

Thayananthan (2005) proposed the Multivariate Relevance Vector Machine (MVRVM) to provide a regression tool capable of generating multivariate outputs. This model is an extension of the sparse Bayesian model developed by Tipping and Faul (2003). It is developed as follows.

Given a training data set of input-target vector pairs  $\{\mathbf{x}^{(n)}, \mathbf{t}^{(n)}\}_{n=1}^N$ , where  $N$  is the number of observations,  $\mathbf{x} \in \mathbb{R}^D$  is a  $D$ -dimensional input vector,  $\mathbf{t} \in \mathbb{R}^M$  is a  $M$ -dimensional output target vector; the model learns the dependency between input and output target with the purpose of making accurate predictions of  $\mathbf{t}$  for previously unseen values of  $\mathbf{x}$ :

$$\mathbf{t} = \mathbf{W} \boldsymbol{\Phi}(\mathbf{x}) + \boldsymbol{\varepsilon} \quad (4.1)$$

where  $\mathbf{W}$  is a  $M \times P$  weight matrix and  $P = N+1$ . The error  $\boldsymbol{\varepsilon}$  is conventionally assumed to be zero-mean Gaussian with diagonal covariance matrix  $\mathbf{S} = \text{diag}(\sigma_1^2, \dots, \sigma_M^2)$ .

A fixed kernel function  $K(\mathbf{x}, \mathbf{x}_n)$  is used to create a vector of basis functions of the form  $\boldsymbol{\Phi}(\mathbf{x}) = [1, K(\mathbf{x}, \mathbf{x}^{(1)}), \dots, K(\mathbf{x}, \mathbf{x}^{(N)})]$ . Tipping (2001) pointed out that this kernel function is used simply to define a set of basis functions, rather than as a definition of a dot product in some feature space such as is employed by the SVM approach. In this paper, we considered a Gaussian kernel  $K(\mathbf{x}, \mathbf{x}_n) = \exp(-r^{-2} \|\mathbf{x} - \mathbf{x}^{(n)}\|^2)$  where  $r$  is called the kernel width parameter, which is a smoothing parameter to define a basis function to capture patterns in the data (Appendix A). This type of kernel has been used by several authors in water resources and hydrology applications (Khalil *et al.* 2005b; Tripathi and Govindaraju 2006).

Let  $\mathbf{t} = [\tau_1, \dots, \tau_r, \dots, \tau_M]^T$  and  $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_r, \dots, \mathbf{w}_M]^T$ . A likelihood distribution of the weight matrix can be written as a product of Gaussians of the weight vectors ( $\mathbf{w}_r$ ) corresponding to each target output ( $\tau_r$ ) (Thayananthan *et al.*, 2008):

$$p(\{\mathbf{t}^{(n)}\}_{n=1}^N | \mathbf{W}, \mathbf{S}) = \prod_{n=1}^N N(\mathbf{t}^{(n)} | \mathbf{W} \Phi(\mathbf{x}^{(n)}), \mathbf{S}) = \prod_{r=1}^M N(\tau_r | \mathbf{w}_r \Phi, \sigma_r^2) \quad (4.2)$$

where  $\Phi = [1, \Phi(\mathbf{x}_1), \Phi(\mathbf{x}_2), \dots, \Phi(\mathbf{x}_N)]$ . Eq. 4.2 therefore contains the parameters  $\mathbf{W}$  (which is  $M$  by  $N$ ) and  $\sigma_r^2$  (which is a vector of length  $M$ ). As a result, there is a danger that the maximum likelihood estimation of  $\mathbf{w}_r$  and  $\sigma_r^2$  will suffer from severe over-fitting. To avoid this, Tipping (2001) proposed constraining the selection of parameters by applying a Bayesian perspective and defining an explicit zero-mean Gaussian prior probability distribution over them (Thayananthan *et al.*, 2008):

$$p(\mathbf{W} | \mathbf{A}) = \prod_{r=1}^M \prod_{j=1}^P N(w_{rj} | 0, \alpha_j^{-2}) = \prod_{r=1}^M N(\mathbf{w}_r | 0, \mathbf{A}) \quad (4.3)$$

where  $\mathbf{A} = \text{diag}(\alpha_1^{-2}, \dots, \alpha_P^{-2})^T$  is a diagonal matrix of hyperparameters  $\alpha_j$ , and  $w_{rj}$  is the  $(r,j)^{\text{th}}$  element of the weight matrix  $\mathbf{W}$ . Each  $\alpha_j$  controls the strength of the prior over its associated weight (Tipping and Faul, 2003).

Bayesian inference considers the posterior distribution of the model parameters, which is proportional to the product of the likelihood and prior distributions:

$$p(\mathbf{W} | \{\mathbf{t}\}_{n=1}^N, \mathbf{S}, \mathbf{A}) \propto p(\{\mathbf{t}\}_{n=1}^N | \mathbf{W}, \mathbf{S}) p(\mathbf{W} | \mathbf{A}) \quad (4.4)$$

The posterior parameter distribution conditioned on the data can be written as the product of Gaussians for the weight vectors of each target output dimension (Thayananthan *et al.*, 2008):

$$p(\mathbf{W} | \{\mathbf{t}\}_{n=1}^N, \mathbf{S}, \mathbf{A}) \propto p(\{\mathbf{t}\}_{n=1}^N | \mathbf{W}, \mathbf{S}) p(\mathbf{W} | \mathbf{A}) \propto \prod_{r=1}^M N(\mathbf{w}_r | \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r) \quad (4.5)$$



The posterior distribution of the weights is Gaussian  $N(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r)$  with the covariance and mean,  $\boldsymbol{\Sigma}_r = (\mathbf{A} + \boldsymbol{\sigma}_r^{-2} \boldsymbol{\Phi}^T \boldsymbol{\Phi})^{-1}$  and  $\boldsymbol{\mu}_r = \boldsymbol{\sigma}_r^{-2} \boldsymbol{\Sigma}_r \boldsymbol{\Phi}^T \boldsymbol{\tau}_r$ , respectively. Given this posterior, we can obtain an optimal weight matrix by estimating a set of hyperparameters that maximizes the data likelihood over the weights in Equation (4.5). The marginal likelihood is then:

$$\begin{aligned} p(\{\mathbf{t}\}_{n=1}^N | \mathbf{A}, \mathbf{S}) &= \int p(\{\mathbf{t}\}_{n=1}^N | \mathbf{W}, \mathbf{S}) p(\mathbf{W} | \mathbf{A}) d\mathbf{W}, \\ &= \prod_{r=1}^M \int N(\boldsymbol{\tau}_r | \mathbf{w}_r \boldsymbol{\Phi}, \sigma_r^2) N(\mathbf{w} | \mathbf{0}, \mathbf{A}) = \prod_{r=1}^M |\mathbf{H}_r|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \boldsymbol{\tau}_r^T \mathbf{H}_r^{-1} \boldsymbol{\tau}_r\right) \end{aligned} \quad (4.6)$$

where  $\mathbf{H}_r = \sigma_r^{-2} \mathbf{I} + \boldsymbol{\Phi} \mathbf{A}^{-1} \boldsymbol{\Phi}^T$ . Then, we can obtain an optimal set of hyperparameters  $\boldsymbol{\alpha}^{\text{opt}} = \{\boldsymbol{\alpha}_j^{\text{opt}}\}_{j=1}^P$  and noise parameters  $(\boldsymbol{\sigma}^{\text{opt}})^2 = \{\sigma_r^{\text{opt}}\}_{r=1}^M$  by maximizing the marginal likelihood using the fast marginal likelihood maximization algorithm proposed by Tipping and Faul (2003). During the optimization process, many elements of  $\boldsymbol{\alpha}$  go to infinity, for which the corresponding posterior probability of the weight becomes zero. The relatively few nonzero weights correspond to the input vectors that form the sparse core of the RVM model. It is these input vectors, called the relevance vectors (RVs), that generate a sparse representation. The optimal parameters are used to obtain the optimal weight matrix with optimal covariance  $\boldsymbol{\Sigma}^{\text{opt}} = \{\boldsymbol{\Sigma}_r^{\text{opt}}\}_{r=1}^M$  and mean  $\boldsymbol{\mu}^{\text{opt}} = \{\boldsymbol{\mu}_r^{\text{opt}}\}_{r=1}^M$ .

We can compute the predictive distribution for any new input  $\mathbf{x}^*$ , corresponding to a target  $\mathbf{t}^*$  (Tipping, 2001), from:

$$p(\mathbf{t}^* | \mathbf{t}, \boldsymbol{\alpha}^{\text{opt}}, (\boldsymbol{\sigma}^{\text{opt}})^2) = \int p(\mathbf{t}^* | \mathbf{W}, (\boldsymbol{\sigma}^{\text{opt}})^2) p(\mathbf{W} | \mathbf{t}, \boldsymbol{\alpha}^{\text{opt}}, (\boldsymbol{\sigma}^{\text{opt}})^2) d\mathbf{W} \quad (4.7)$$

Taking into consideration that both terms in the integrand are Gaussian, Equation (4.7) is computed as:

$$p(\mathbf{t}^* | \mathbf{t}, \mathbf{a}^{\text{opt}}, (\boldsymbol{\sigma}^{\text{opt}})^2) = \mathbf{N}(\mathbf{t}^* | \mathbf{y}^*, (\boldsymbol{\sigma}^*)^2) \quad (4.8)$$

where  $\mathbf{y}^* = [y_1^*, \dots, y_r^*, \dots, y_M^*]^T$  is the predictive mean with  $y_r^* = (\boldsymbol{\mu}_r^{\text{opt}})^T \boldsymbol{\Phi}(\mathbf{x}^*)$ ; and  $(\boldsymbol{\sigma}^*)^2 = [(\sigma_1^*)^2, \dots, (\sigma_r^*)^2, \dots, (\sigma_M^*)^2]^T$  is the predictive variance with  $(\sigma_r^*)^2 = (\sigma_r^{\text{opt}})^2 + \boldsymbol{\Phi}(\mathbf{x}^*)^T \boldsymbol{\Sigma}_r^{\text{opt}} \boldsymbol{\Phi}(\mathbf{x}^*)$ .  $(\sigma_r^*)^2$  contains the sum of two variance terms: the noise on the data and the uncertainty in the prediction of the weight parameters (Tipping, 2001).

The standard deviation  $\sigma_r^*$  of the predictive distribution is defined as a predictive error bar of  $y_r^*$  (Bishop, 1995). Then, the width of the 90% predictive confidence interval for any  $y_r^*$  can be calculated as  $\pm 1.65 \sigma_r^*$ .

Readers interested in greater detail regarding multivariate sparse Bayesian regression, its mathematical formulation, and the optimization procedures of the model are referred to Thayananthan (2005), Thayananthan *et al.* (2008), Tipping (2001), and Tipping and Faul (2003). A MATLAB code developed by Thayananthan (2005) is available from <http://mi.eng.cam.ac.uk/~at315/MVRVM>.

#### 4.2.2 Study area and data

The Sevier River Basin is used here to demonstrate the MVRVM modeling approach as applied to a river reach located between two reservoirs that are in series. The basin is located in south-central Utah and is the largest drainage area in the state (approximately 12.5 percent of the state's area). Average annual precipitation ranges from 6.4 to 13.0 inches in the valleys to more than 40 inches in the high mountains. Elevation, precipitation, and temperatures are highly variable over the basin, and as a result there are several vegetative types that grow in the area. The main use of water in the basin is for irrigation. The economy of the Basin is based primarily on agriculture, tourism, and a few mining and manufacturing enterprises (Berger *et al.*, 2003).

In this paper we focus on a reach of Sevier River which is regulated by Sevier Bridge (or Yuba) reservoir to fulfill irrigation canal orders (Central Utah Canal, Vincent Canal and Leamington Canal) and also provide water to the Delta-Millard Association Dam (DMAD) Reservoir (Figure 4.1). This study river reach is between Sevier Bridge reservoir and Sevier River near Lynndyl station (Figure 4.1).

A seepage study of the Sevier River and irrigation canals was carried out by the United State Geological Survey, USGS (1981). This report showed a net gain of 9 cfs in the river reach that is the object of study in this paper, a net loss of 7 cfs in the Central Utah Canal, a net loss of 0.8 cfs in the Vincent Canal, and a net gain of about 1.3 cfs in the Leamington Canal. This report pointed out that losses from the Vincent Canal may contribute to the gains in the Sevier River. Also, losses from the Central Utah Canal may contribute to the gains in the Sevier River and Leamington Canal; and additional gains in the Sevier River are probably inflows from unconsumed irrigation water seeping into permeable unconsolidated deposits through which it moves to the river. In fact, Frevert (1982) indicated that return flow into the study reach comes primarily from springs immediately downstream of Sevier Bridge reservoir and return flow from the irrigation canals mentioned above.

Figure 4.2a shows the daily streamflow variation on the Sevier River near Lynndyl station during the 2007 irrigation season. The calculated daily water balance in the study river reach during the 2007 irrigation season showed unpredictable variation due to water loss/gain in the river reach (Figure 4.2b). This water balance calculation will be explained in the next subsection.

#### 4.2.3 Identification of inputs and outputs

Monitoring data are posted hourly to the Sevier River Water Users website, [www.sevierriver.org](http://www.sevierriver.org). This information, in combination with internet-based controls, enables real-time operations of reservoir releases and canal diversions throughout the entire river basin. Daily averages taken from the Sevier River database from 2001 to 2007 were used to build the MVRVM model. Daily data from the irrigation seasons of 2001 through 2006 were used to train the MVRVM and find the model parameters. Daily data from the 2007 irrigation season were used to test the model.

The gauging station located on the Sevier River near Juab measures the releases from Sevier Bridge Reservoir into the Sevier River. Four gauging stations are located in the river reach between Sevier Bridge and DMAD reservoirs: three measure diversions to irrigation canals, and one station measures streamflow on the Sevier River near Lynndyl (Figure 4.1). The river reach that is the object of this study is located between the station near Juab and that near Lynndyl.

Average travel times during the irrigation season were taken into consideration in order to lag the records at Sevier River near Lynndyl through the study reach and estimate the net cumulative water loss/gain. These approximated daily travel times were estimated according to information provided by the reservoir operator (unpublished data, 2007). The average lag time between the starting time of the releases at Sevier Bridge reservoir and when the water reaches the station near Lynndyl is approximately two days. The average lag time between the time when the streamflow reaches the irrigation canal diversion stations and when the water reaches the Lynndyl station is one day, approximately.

Based on the average travel times described above, the cumulative net water balance on any day “d” can be estimated in the study river reach:

$$\mathbf{LG}_d = \mathbf{SRNL}_d - (\mathbf{SB}_{d-2} - \mathbf{CUC}_{d-1} - \mathbf{VC}_{d-1} - \mathbf{LC}_{d-1}) \quad (4.9)$$

where,

$\mathbf{LG}_d$  = Cumulative net water loss (negative) or gain (positive) volume in the river reach on any day “d” (acre feet)

$\mathbf{SRNL}_d$  = Cumulative volume at Sevier River near Lynndyl on any day “d” (acre feet)

$\mathbf{SB}_{d-2}$  = Daily cumulative volume from Sevier Bridge Reservoir two days prior to day “d” (acre feet)

$\mathbf{CUC}_{d-1}$  = Daily cumulative volume diverted into Central Utah canal one day prior to day “d” (acre feet)

$\mathbf{VC}_{d-1}$  = Daily cumulative volume diverted into Vincent Utah canal one day prior to day “d” (acre feet)

$\mathbf{LC}_{d-1}$  = Daily cumulative volume diverted into Leamington Utah canal one day prior to day “d” (acre feet)

The cumulative net water balance over a two day period can be estimated as follows:

$$\mathbf{LG}_\Sigma = \mathbf{LG}_d + \mathbf{LG}_{d+1} \quad (4.10)$$

where,

$\mathbf{LG}_\Sigma$  = Cumulative net water loss/gain in the river reach over a two day period (acre feet)

$\mathbf{LG}_d$  = Cumulative net water loss/gain in the river reach one day ahead (acre feet)

$\mathbf{LG}_{d+1}$  = Cumulative net water loss/gain in the river reach two days ahead (acre feet)

Air temperature can directly influence some hydrological cycle processes such as evapotranspiration in the irrigated areas and evaporation in the river. These processes can

influence the daily variability of streamflow and water loss/gain in the river reach. Daily maximum and minimum temperature from Oak City was included in the model inputs. This historical data was obtained from The National Oceanic and Atmospheric Administration (NOAA).

The inputs used in the model are expressed as:

$$\mathbf{x} = [\mathbf{DC}_{d-nd}, \mathbf{SF}_{d-nd}, \mathbf{SB}_{d-nd}, \mathbf{T}_{d-nd}, \mathbf{LG}_{d-nd}]^T \quad (4.11)$$

where,

$d$  = day of prediction

$nd$  = number of days previous to the prediction time

$\mathbf{DC}_{d-nd}$  = diversions to the Central Utah canal, Vincent canal, Leamington canal.

$\mathbf{SF}_{d-nd}$  = streamflow on the Sevier River near Lynndyl.

$\mathbf{SB}_{d-nd}$  = Sevier Bridge reservoir releases.

$\mathbf{T}_{d-nd}$  = maximum and minimum daily temperature from Oak City.

$\mathbf{LG}_{d-nd}$  = Cumulative net water loss/gain over a one-day period (acre feet)

The multiple output target vector of the model is expressed as

$$\mathbf{t} = [\mathbf{SF}_d, \mathbf{SF}_{d+1}, \mathbf{LG}_d, \mathbf{LG}_\Sigma]^T \quad (4.12)$$

where,

$\mathbf{SF}_d$  = prediction of streamflow on the Sevier River near Lynndyl one day ahead (cfs)

$\mathbf{SF}_{d+1}$  = prediction of streamflow on the Sevier River near Lynndyl two days ahead (cfs)

$\mathbf{LG}_d$  = prediction of the cumulative net water loss/gain over a one-day period (acre feet per day)

$\mathbf{LG}_\Sigma$  = prediction of the cumulative net water loss/gain over a two-day period (acre feet)

Finally, the model can be defined as in Equation (4.1) with a data set of input-output pairs  $\{\mathbf{x}^{(n)}, \mathbf{t}^{(n)}\}_{n=1}^N$ , where  $N$  is the number of observations.

#### 4.2.4 Model selection

The statistic used for model selection is the coefficient of efficiency ( $E$ ) for the testing phase. This has been recommended by the ASCE (1993) and Legates and McCabe (1999), and is given by:

$$E = 1 - \frac{\sum_{n=1}^N (t_n - t_n^*)^2}{\sum_{n=1}^N (t_n - t_{av})^2} \quad (13)$$

where  $t$  is the observed output;  $t^*$  is the predicted output;  $t_{av}$  is the observed average output; and  $N$  is the number of observations. This statistic ranges from minus infinity (poor model) to 1.0 (a perfect model) (Legates and McCabe, 1999).

The kernel width parameter had to be selected in order to find the appropriated model (Tipping, 2001; Ghosh and Mujumdar, 2007). Several MVRVM models were built with variation in kernel width and the number of previous days of historical data as inputs. The number of days previous to the prediction time ranges from 1 to 5 days. For each previous day the kernel width ranges from 1 to 40. The selected kernel width is the one with maximum  $E$ . From the list of models with selected kernel width at different previous days, we consider that the best model is the one with the maximum  $E$  and minimum number of days as input data.

#### 4.2.5 Bootstrap analysis

Hydrologic processes are intrinsically nonlinear and generally show high variability in space and time. This is due to variability in climatic inputs and

heterogeneities in catchments and landscape properties (Sivakumar, 2009). In this paper, we apply machine learning techniques for hydrological modeling to learn inverse or retrieval models (Cherkassky *et al.*, 2006; Krasnopolsky, 2009). This means we are training a model to find its parameters based on measured data. Most geophysical and hydrological inverse processes are ill-posed (Parker, 1994), so machine learning applications for hydrologic process deal with formulation for ill-posed problems while trying to capture the underlying nonlinear functions of dynamic systems based on measured data. Also, there are uncertainties in data measurements (e.g. noise in data, scarce data, lack of relevant data, etc.) which contribute to data complexity (Cherkassky *et al.*, 2006). This is why it is necessary to develop machine learning models that guaranty high accuracy and that possess good generalization and robustness characteristics with respect to future changes in the nature of the input data.

The bootstrap method was used to explore the implications of the change in the nature of input data and to guarantee good generalization ability and robustness of the machine learning model (Khalil *et al.*, 2005b). The bootstrap data set was created by randomly selecting from the whole training data set, with replacement. In this bootstrap estimation, the selection process was independently repeated 1,000 times to yield 1,000 bootstrap training data sets, which are treated as independent sets (Duda *et al.*, 2001). For each of the bootstrap training data sets, a model was built and evaluated over the original test data set.

Efron and Tibshirani (1998) emphasized that it is always wise to look at the bootstrap data graphically, rather than relying entirely on a single summary statistic estimator. The bootstrap method provides information on the uncertainty in the statistics



estimator evaluated in the model. Khalil *et al.* (2005b) pointed out that a robust model is one that shows narrow confidence bounds in the bootstrap histogram.

#### 4.2.6 Comparison between MVRVM and ANN

ANNs have been widely applied in irrigation and water resources and hydrologic modeling (ASCE Task Committee on the Application of ANNs in Hydrology, 2000a, 2000b; Khalil *et al.*, 2005d; Pulido-Calvo and Gutierrez-Estrada, 2008; Khoob, 2008; Kisi, 2009; Adedoye, 2009). A comparative analysis between the developed MVRVM models and ANN models is performed to contrast performance and robustness. Readers interested in greater detail regarding ANNs and their training functions are referred to Demuth *et al.* (2009).

The ANN toolbox available in MATLAB is applied in this research. The Levenberg-Marquardt (LM) algorithm was used to train the ANN model. This is a training algorithm with second-order convergence to optimize the ANN parameters (Tan and Cauwenbergh, 1999) and it has been recommended by several authors in irrigation, water resources, and hydrologic research (Anctil *et al.*, 2003; Anctil and Rat, 2005; Pulido-Calvo and Gutierrez-Estrada, 2008; Khoob, 2008; Kisi, 2009).

Several feed-forward ANN models were built with variation in size of the hidden layer and the number of previous time steps to include in the training data. The number of days previous to the prediction time required as input ranges from 1 to 5 days. For each set of previous days, the size of the hidden layer ranges from 1 to 10. The selected layer size is the one with maximum E. Similar to the MVRVM model selection and having the

list of selected models at different previous days as input, the best model is the one with the maximum E and minimum number of previous days.

#### 4.3 Results and Discussion

In the RVM approach, the relevance vectors (RVs) are subsets of the training data set that are used for prediction (Khalil *et al.*, 2005b); as a consequence, the complexity of the model is proportional to the number of RVs. The model only utilizes 28 RVs from the full data set (1182 observations) that was used for training (2001 through 2006 irrigation seasons). This small number of vectors illustrates that the Bayesian learning procedure embodied in the MVRVM is capable of producing very sparse models.

The predicted outputs of the MVRVM for the testing phase (2007 irrigation season) are shown as the solid lines in Figure 4.3. The model explains well the observed streamflows and water loss/gain (dots). The performance accuracy is reduced for the cumulative water loss/gain over two days (Figure 4.3d). This accuracy reduction is found in most time series models, where the farther we predict into the future, the less accurate the prediction becomes. Figure 4.3 also shows the 0.90 confidence interval (shaded region) associated with the predictive variance of the MVRVM in Equation (8).

The observed (dots) and predicted (solid lines) outputs of the ANN for the testing phase (2007 irrigation season) are shown Figure 4.4. The models explain well the observed streamflows and water loss/gain (dots). As with the MVRVM model, the performance accuracy is reduced for the cumulative water loss/gain over two days (Figure 4.4d).

Table 4.1 shows some statistics that measure MVRVM and ANN performance. Again, we can see good performance of both models in the testing phase. The

performance accuracy is reduced for the cumulative water loss/gain over two days. It is important to mention that the selected model for the MVRVM needed two days as input data, and the selected model for the ANN needed four days as input data. We can see that the performance results for MRVM and ANN are fairly similar.

Figures 4.5 and 4.6 show the bootstrap histograms for the RMSE test based on 1,000 bootstrap training data sets of the MVRVM and ANN models. The width of the bootstrapping confidence intervals provides information on the uncertainty in the model parameters. A narrow confidence interval implies low variability of the statistics with respect to possible future changes in the input data, which indicates that the model parameter set is robust [Khalil *et al.*, 2005c]. The bootstrap histograms of the MVRVM model (Figure 4.5) show very narrow confidence bounds in comparison to the histograms of the ANN model (Figure 4.6). Therefore, the MVRVM allows more robust performance.

Fig 4.7 shows a narrow confidence interval on the bootstrap histogram of the number of RVs used in the MVRVM model. The low variability of the number of RVs indicates that the model structure is stable and robust with respect to future changes in the nature of the input data.

#### 4.4 Summary and Conclusions

This paper presents a first attempt to use a machine learning model to develop daily multiple predictions of two non-linear hydrological processes: streamflow and water loss/gain in a river reach. These processes have high variability in space and time due to variability in climatic inputs and heterogeneities in catchments and landscape properties. A physical model may require detailed input information that maybe too costly

or impractical to obtain for real-time applications. That is why we propose the application of the MVRVM, which is a Bayesian machine learning regression tool extension of the RVM algorithm to produce multivariate probabilistic outputs (with predictive confidence intervals) when given a set of multivariate inputs. These inputs are commonly available past data (releases from upstream reservoir, streamflow, river outflows and air temperature) and the outputs are streamflow one and two days into the future, and net cumulative water loss/gain in a river reach over the same two-day period. The model is illustrated by application to a river system located at the Lower Sevier River near Delta, Utah. The results show that the model learns the patterns between outputs and inputs for the training phase and makes accurately predictions for the testing phase.

The MVRVM model has the property of sparse formulation. The parsimonious structure of this empirical model is sufficient to explain the data and to avoid data overfitting. Therefore, we can see an important advantage of the Bayesian learning procedure, which is the capability of the MVRVM to produce very sparse models. Besides this sparse formulation, an important advantage of utilizing MVRVM is its generalization capabilities to predict future system states. The dynamics of streamflow and water loss/gain processes could become difficult to predict since they are nonlinear and mostly show high variability in space and time. The model presented here ensures good generalization providing robustness with respect to new input data.

The performance results are fairly similar for the MVRVM and ANN models. However, the bootstrap histograms of the MVRVM model show narrow confidence bounds in comparison to the histograms of the ANN model. Therefore, the MVRVM is more robust.

In summary, the results presented in this paper have demonstrated the successful performance and robustness of MVRVM in predicting the non-linear behavior of hydrological processes in a river system requiring fewer inputs than physical models. Simultaneous multiple-time-ahead streamflow and water loss/gain predictions from a river system have potential value to assist the reservoir operator in efficiently selecting the real-time operation and management decisions for river systems.

## References

- Adeloye AJ. 2009. Multiple linear regression and artificial neural networks models for generalized reservoir storage-yield-reliability function for reservoir planning. *Journal of Hydrologic Engineering* **14**: 731-738. DOI: 10.1061/(ASCE)HE.1943-5584.0000041
- Anctil F, Perrin C, Andréassian V. 2003. ANN output updating of lumped conceptual rainfall/runoff forecasting models. *Journal of American Water Resources Association* **35**: 1269–1279. DOI: 10.1111/j.1752-1688.2003.tb03708.x
- Anctil F, Rat A. 2005. Evaluation of neural network streamflow forecasting on 47 watersheds. *Journal of Hydrologic Engineering* **10**: 85–88. DOI: 10.1061/(ASCE)1084-0699(2005)10:1(85)
- ASCE Task Committee on Definition of Criteria for Evaluation of Watershed Models of the Watershed Management, Irrigation, and Drainage Division (ASCE). 1993. Criteria for evaluation of watershed models. *Journal of Irrigation and Drainage Engineering* **119**: 429-442. DOI: 10.1061/(ASCE)0733-9437(1993)119:3(429)
- ASCE Task Committee on the Application of ANNs in Hydrology. 2000a. Artificial neural networks in hydrology, I: Preliminary concepts. *Journal of Hydrologic Engineering* **5**: 115-123. DOI: 10.1061/(ASCE)1084-0699(2000)5:2(115)
- ASCE Task Committee on the Application of ANNs in Hydrology. 2000b. Artificial neural networks in hydrology, II: Hydrologic application. *Journal of Hydrologic Engineering* **5**: 124-137. DOI: 10.1061/(ASCE)1084-0699(2000)5:2(124)
- Berger B, Hansen R, Jensen R. 2003. Sevier River Basin System Description, Sevier River Water Users Association, Delta, Utah.

- Birikundavyi S, Labib R, Trung H, Rousselle J. 2002. Performance of neural networks in daily streamflow forecasting. *Journal of Hydrology Engineering* **7**: 392-398. DOI: 10.1061/(ASCE)1084-0699(2002)7:5(392)
- Bishop C M. 1995. *Neural Network for Pattern Recognition*. Oxford University Press: New York.
- Cherkassky V, Krasnopolsky V, Solomatine DP, Valdes J. 2006. Computational intelligence in earth sciences and environmental applications: issues and challenges. *Neural Networks* **19**:113–121. DOI: 10.1016/j.neunet.2006.01.001
- Demuth, H, Beale M, Hagan M. 2009. *Neural Network Toolbox user's guide*, The MathWorks Inc, MA, USA.
- Duda RO, Hart P, Stork D. 2001. *Pattern Classification*. Wiley Interscience, Second Edition, New York.
- Efron B, Tibshirani R. 1998. *An Introduction of the Bootstrap*, Monographs on Statistics and Applied Probability 57, CRC Press LLC, Florida.
- Frevert DK. 1983. Prediction of irrigation return flow. *PhD Dissertation*, Agricultural and Chemical Engineering Department, Colorado State University.
- Ghosh S, Mujumdar PP. 2007. Statistical downscaling of GCM simulations to streamflow using relevance vector machine. *Advances in Water Resources* **31**: 132-146. DOI:10.1016/j.advwatres.2007.07.005
- Hurkmans RTWL, de Moel H, Aerts JC JH, Troch PA. 2008. Water balance versus land surface model in the simulation of Rhine river discharges. *Water Resources Research* **44**:W01418. DOI: 10.1029/2007WR006168
- Ivkovic KM. 2009. A top-down approach to characterize aquifer-river interaction processes. *Journal of Hydrology* **365**: 145-155. DOI: 10.1016/j.jhydrol.2008.11.021
- Jothiprakash V. 2003. Water balance model to assess the water loss/gain in a river system. *Journal of Institution of Engineers (India)* **84**:196 – 200.
- Krasnopolsky VM. 2009. Neural network applications to solve forward and inverse problems in atmospheric and oceanic satellite remote sensing. *Artificial Intelligence Methods in the Environmental Sciences* **9**: 191-205. DOI: 10.1007/978-1-4020-9119-3\_9
- Khalil A, Almasari M, McKee M, Kemblowski MW, Kaluarachchi J. 2005a. Applicability of statistical learning algorithms in groundwater quality modeling, *Water Resources Research* **41**:W05010. DOI: 10.1029/2004WR003608

- Khalil A, McKee M, Kemblowski MW, Asefa T. 2005b. Sparse Bayesian learning machine for real-time management of reservoir releases. *Water Resources Research* **41**: W11401. DOI: 10.1029/2004WR003891
- Khalil A, McKee M, Kemblowski MW, Asefa T, Bastidas L. 2005c. Multiobjective analysis of chaotic dynamic systems with sparse learning machines. *Advances in Water Resources* **29**: 72-88. DOI:10.1016/j.advwatres.2005.05.011
- Khalil A, McKee M, Kemblowski MW, and Asefa T. 2005d. Basin-scale water management and forecasting using neural networks, *Journal of American Water Resources Research* **41**: 195-208.
- Khoob AR. 2008. Artificial neural network estimation of reference evapotranspiration from pan evaporation in a semi-arid environment. *Irrigation Science* **27**: 35-39.
- Kisi O. 2007. Streamflow forecasting using different artificial neural network algorithms. *Journal of Hydrologic Engineering* **12**: 532–539. DOI: 10.1061/(ASCE)1084-0699(2007)12:5(532)
- Kisi O. 2009. Neural Networks and Wavelet Conjunction Model for Intermittent Streamflow Forecasting. *Journal of Hydrologic Engineering* **14**: 773-782. DOI:10.1061/(ASCE)HE.1943-5584.0000053
- Legates D R, McCabe GJ. 1999. Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. *Water Resources Research* **35**: 233-241.
- Parker RL. 1994. *Geophysical Inverse Theory*. Princeton University Press. Princeton, NJ.
- Pulido-Calvo I, Gutierrez-Estrada JC. 2008. Improved irrigation water demand forecasting using a soft-computing hybrid model. *Biosystems Engineering* **102**: 202-218. DOI:10.1016/j.biosystemseng.2008.09.032.
- Reddy J M, Wilamowski BM. 2000. Adaptive neural networks in regulation of river flows. *Artificial Neural Network in Hydrology* **8**: 153-177. Netherlands.
- Sivakumar B. 2009. Nonlinear dynamics and chaos in hydrologic systems: latest developments and a look forward. *Stochastic Environmental Research and Risk Assessment* **23**: 1027-1036. DOI: 10.1007/s00477-008-0265-z
- Solomatine DP, Shrestha DL. 2009. A novel method to estimate model uncertainty using machine learning techniques. *Water Resources Research* **45**:W00B11. DOI:10.1029/2008WR006839

- Tan T, Van Cauwenberghe A. 1999. Neural-network-based d-step-ahead predictor for nonlinear systems with time delay. *Engineering Applications of Artificial Intelligence* **12**:21-35.
- Thayananthan A. 2005. Template-based pose estimation and tracking of 3D hand motion. *PhD Thesis*, Department of Engineering, University of Cambridge, Cambridge, United Kingdom.
- Thayananthan A, Navaratnam R, Stenger B, Torr PHS, Cipolla R. 2008. Pose estimation and tracking using multivariate regression. *Pattern Recognition Letters* **29**: 1302-1310.
- Tipping ME. 2001. Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning* **1**: 211–244.
- Tipping M., Faul A. 2003. Fast marginal likelihood maximization for sparse Bayesian models. Paper presented at Ninth International Workshop on Artificial Intelligence and Statistics, Soc. for Artif. Intel. and Stat., Key West, Florida.
- Tripathi S, Govindaraju R. 2006. On selection of kernel parameters in relevance vector machines for hydrologic applications. *Stochastic Environmental Research and Risk Assessment* **21**:747-764. DOI: 10.1007/s00477-006-0087-9
- United State Geological Survey (USGS). 1981. Seepage study of the Sevier River and the Central Utah, McIntyre, and Leamington canals, Juab and Millard counties. All U.S. Government Documents (Utah Regional Depository), Utah.
- Wu JS, Han J, Annambhotla S, and Bryant S. 2005. Artificial neural networks for forecasting watershed runoff and stream flows. *Journal of Hydrologic Engineering* **10**: 216-222. DOI 10.1061/(ASCE)1084-0699(2005)10:3(216)



Table 4.1. Model performance using different statistics

Model	Statistics	Streamflow		Loss/gain	
		$SF_d$	$SF_{d+1}$	$LG_d$	$LG_{d2}$
MVRVM	E	0.99	0.97	0.69	0.55
	RMSE, cfs	23.83	37.43	47.62	97.82
	Predictive error, cfs	22.07	40.53	45.45	90.86
ANN	E	0.99	0.97	0.69	0.58
	RMSE, cfs	24.61	39.45	47.79	94.46

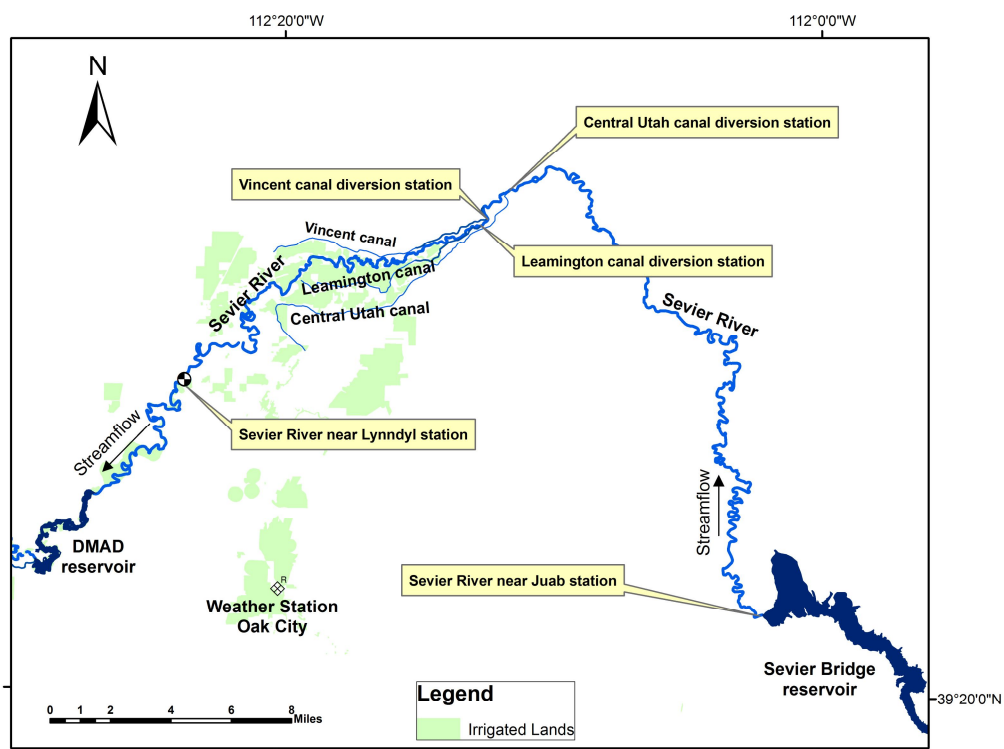


Figure 4.1. Sensor station locations at the study river reach at Lower Sevier River Basin, Utah

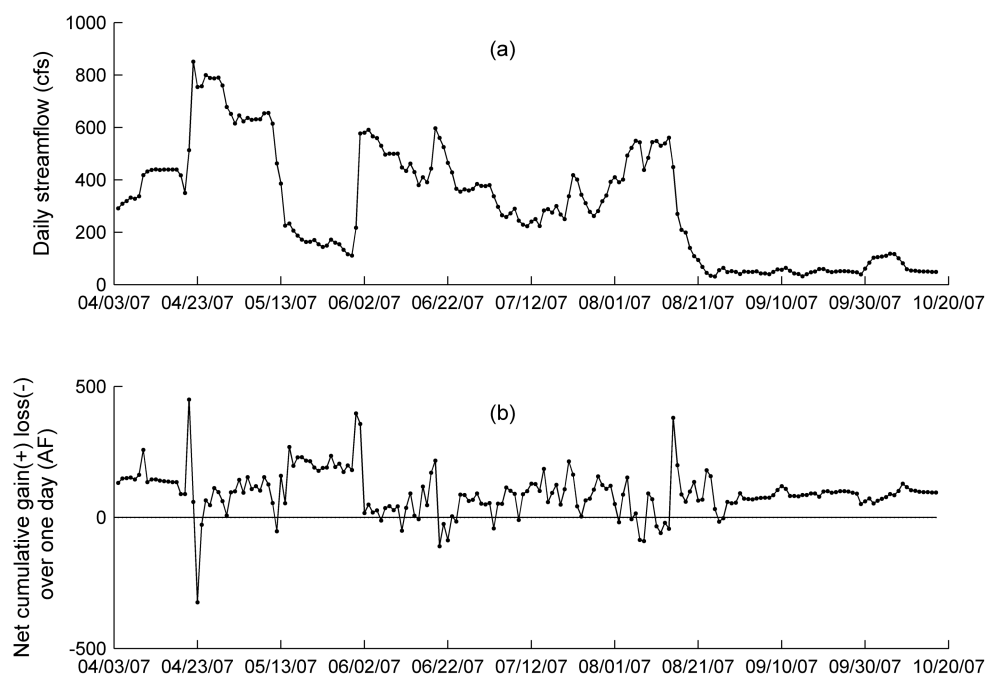


Figure 4.2. (a) Daily streamflow on the Sevier River near Lynndyl, (b) Daily water balance in the study river reach

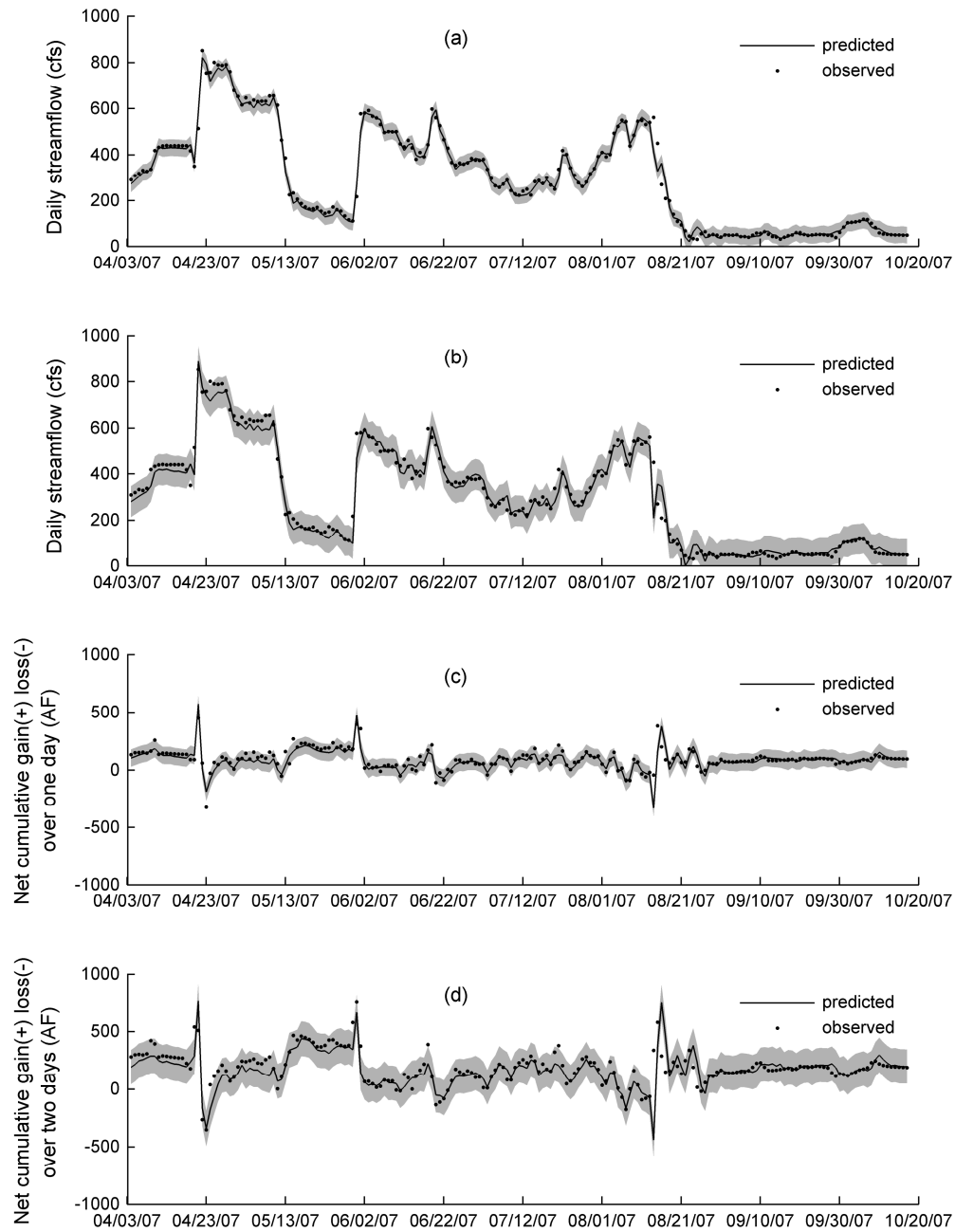


Figure 4.3. Observations versus predictions of the MVRVM with 0.90 confidence intervals (shaded region) for the testing phase (2007 irrigation season): (a) Prediction of streamflow one day ahead, (b) Prediction of streamflow two days ahead, (c) Prediction of net cumulative loss/gain water over one day, (d) Prediction of net cumulative loss/gain water over two days

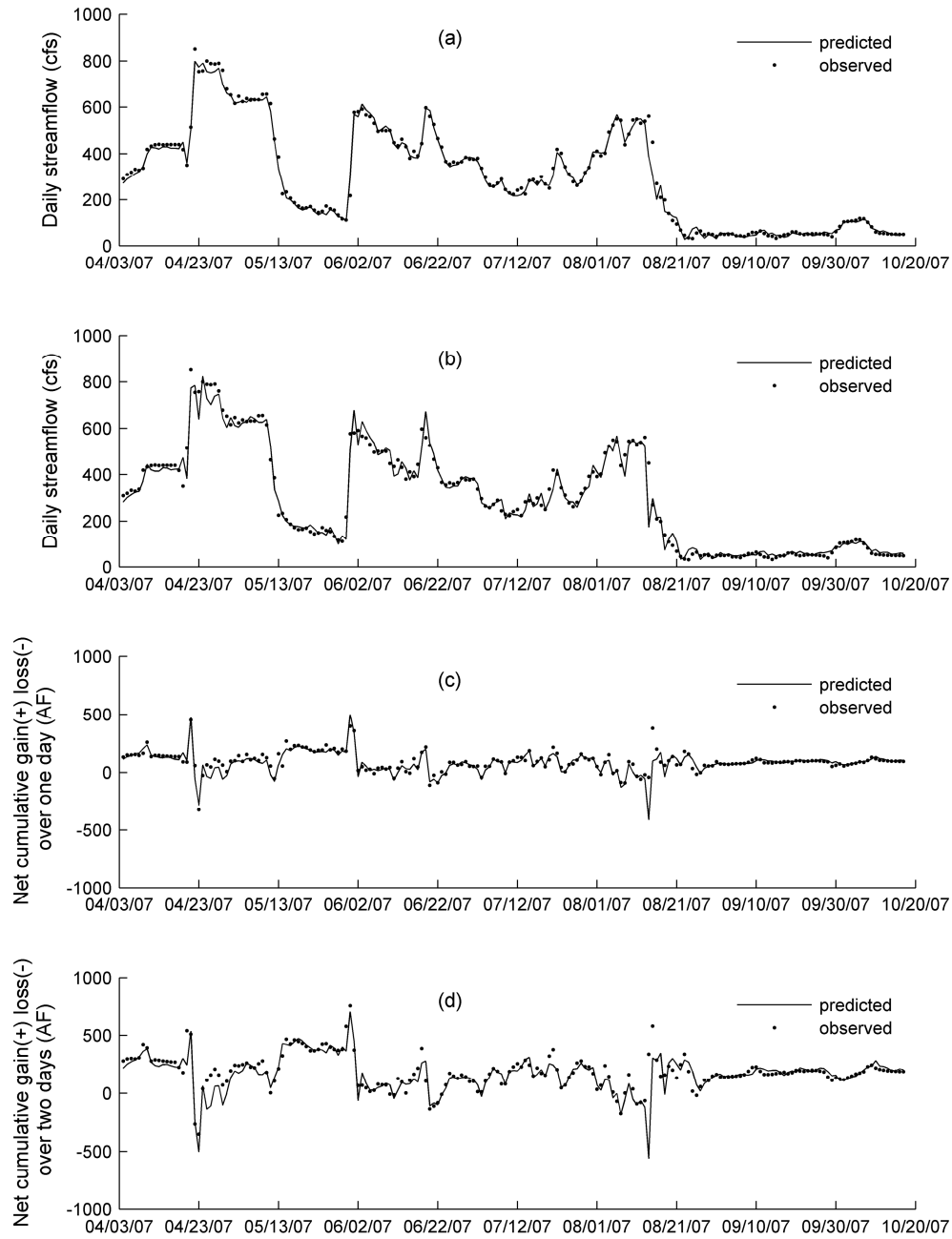


Figure 4.4. Observations versus predictions of the ANN for the testing phase (2007 irrigation season): (a) Prediction of streamflow one day ahead, (b) Prediction of streamflow two days ahead, (c) Prediction of net cumulative loss/gain water over one day, (d) Prediction of net cumulative loss/gain water over two days

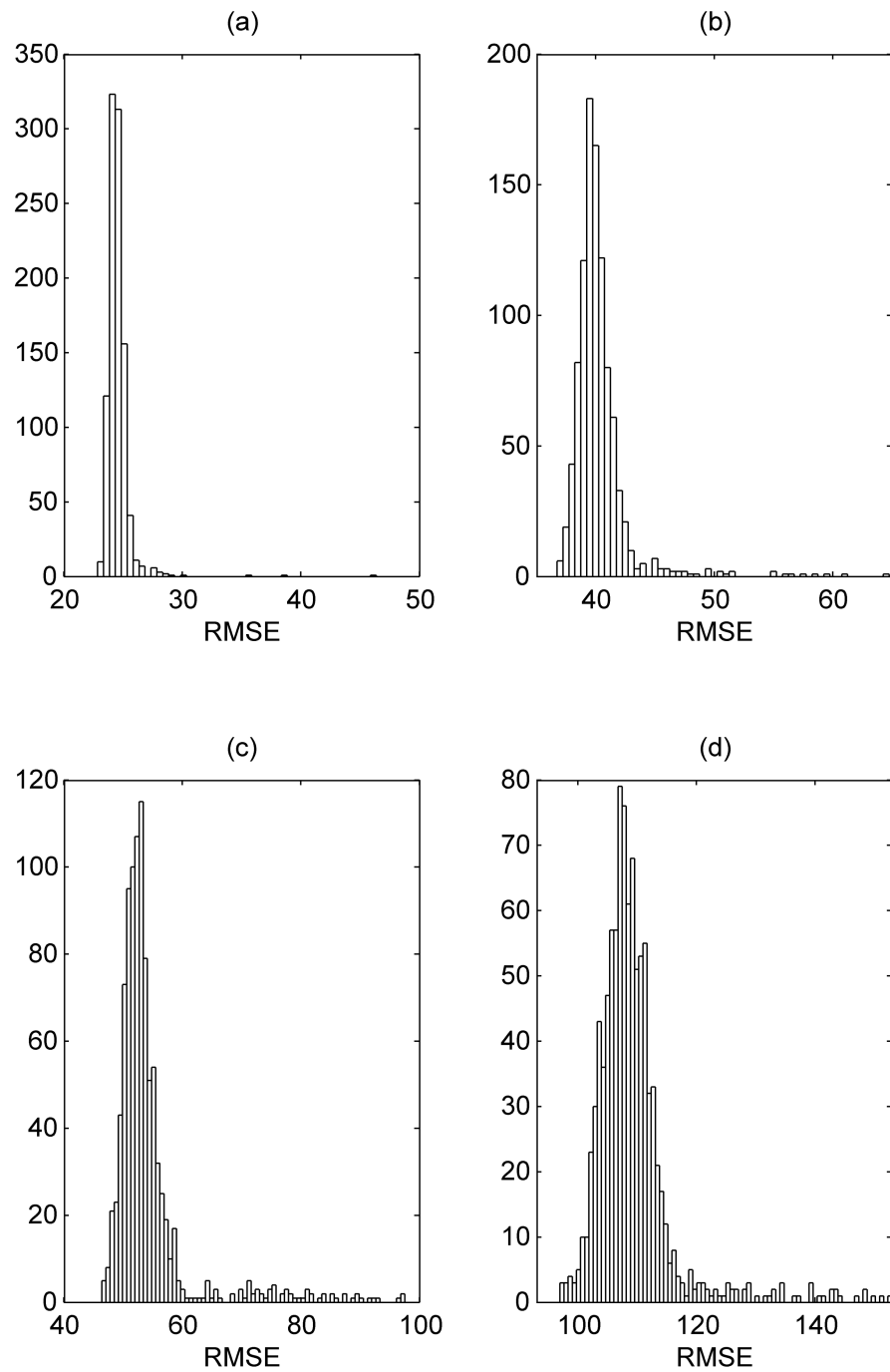


Figure 4.5. Bootstrap histograms of the MVRVM model for the RMSE test. Streamflow predictions: (a) 1-day ahead, (b) 2-days ahead; net cumulative water loss/gain predictions: (c) over one day, (d) over two days

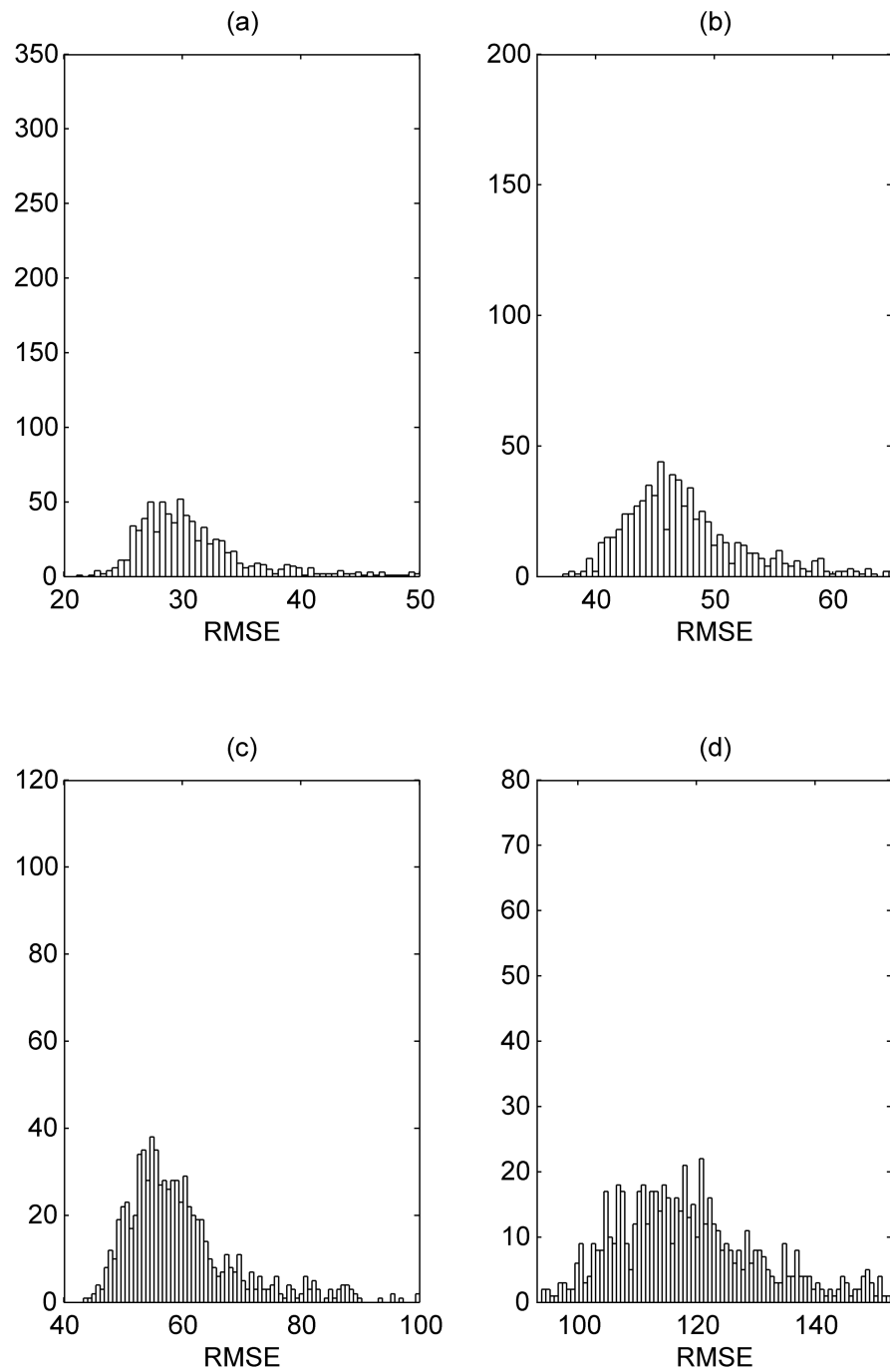
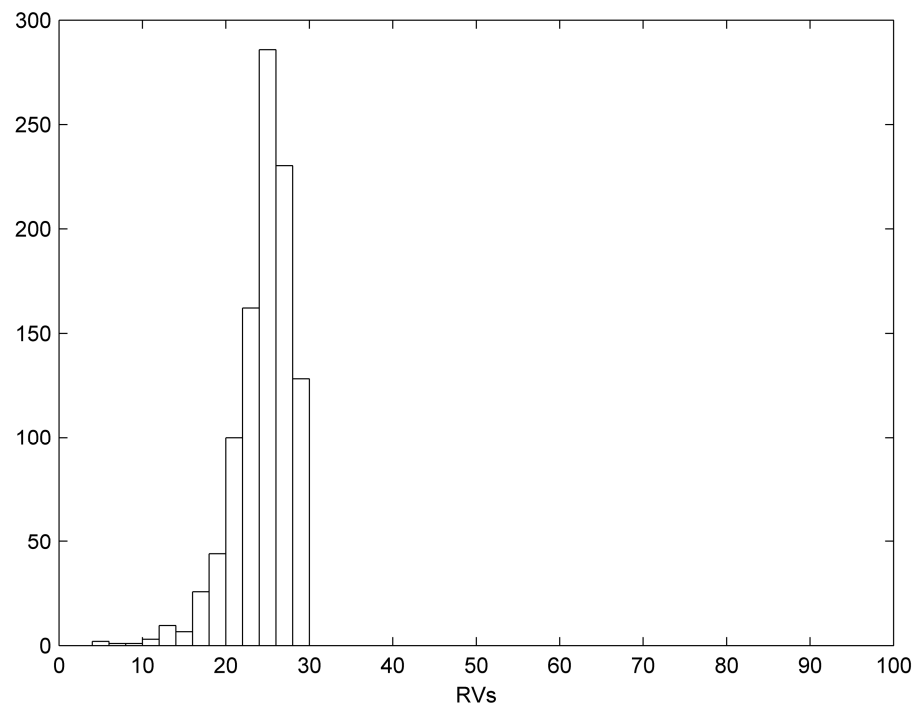


Figure 4.6. Bootstrap histograms of the ANN model for the RMSE test. Streamflow predictions: (a) 1-day ahead, (b) 2-days ahead; net cumulative water loss/gain predictions: (c) over one day, (d) over two days



4.7. Bootstrap histogram of the MVRVM model for the number of RVs



## CHAPTER 5

### SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

#### 5.1 Summary and Conclusions

Application of new multivariate regression techniques for forecasting multiple future system states in a complex regulated river basin system is addressed in this thesis. The focus is to develop a practical, sparse, and robust forecasting model based on machine learning (ML) in order to provide valuable, real-time information to the operators of a complex river basin system in the form of multiple probabilistic predictions (with confidence intervals) of the future states of the water resources system without the application of traditional physically based models.

The Multivariate Relevance Vector Machine (MVRVM) (Thayananthan et al. 2008) model is utilized in this thesis. This dissertation proves that this model, via its Bayesian formulation, can accurately forecast multivariate outputs in a highly complex system (i.e. regulated river basin system). Using this Bayesian approach, a predictive confidence interval is obtained from the model that captures the uncertainty of both the model and the data. The resulting model is parsimonious and robust. The MVRVM models were applied to the multiple reservoir, canal, and river system located in the Lower Sevier River Basin in Utah.

Chapter 2 presents a MVRVM model that simultaneously forecasts water releases one and two days ahead from two reservoirs (Sevier Bridge reservoir and the DMAD reservoir), where the model inputs are available past daily data collected by sensors on the reservoirs, canals, diversions, weather stations, and the river.

In Chapter 3, two MVRVM models are applied to develop multiple-time-ahead predictions of required hourly (one, twelve and twenty-four hours ahead) and daily (one and two days ahead) diversions for three irrigation canals (Central Utah Canal, Vincent Canal and Leamington Canal). The inputs for the hourly model are the past hourly observations of water diversions for the three canals. Inputs for the daily model are the past daily observations of water diversions for three canals and climatic data.

Chapter 4 presents a MVRVM model that simultaneously predicts the non-linear behavior of hydrological processes in a river system, streamflow one and two days into the future, and net cumulative water loss/gain in a river reach over the same two-day period. Water loss/gain process involves return flows, transmission losses, tributary contributions, stream channel characteristics, and other hydrologic factors that are not taken into account in the analysis. Instead, we used a simple water balance from available measured data on the river to estimate the water loss or gain and developed a ML model to learn and mimic the underlying physics by simple examination of inputs and outputs. Inputs to the model are available past data describing releases from the upstream reservoir, streamflow, river outflows, and air temperature. The study river reach is regulated by releases from Sevier Bridge Reservoir to fulfill downstream irrigation canal orders (Central Utah Canal, Vincent Canal and Leamington Canal) and to provide water to the DMAD Reservoir. The reservoir operator can take into account these real-time predictions and decide how to manage releases from the upper reservoir into the lower one.

The hypothesis addressing accuracy and sparse formulation of the MVRVM models can be confirmed by the test results summarized in Table 1. In this table, we can

see the overall average coefficient of efficiency ( $E_{av}$ ) of all the multiple output variables for each model. These efficiencies show that the models perform remarkably well with an  $E_{av}$  ranging from 0.76 to 0.89.

**Table 5.1** Overall performance of the MVRVM models

<b>Multivariate output variables</b>	<b>Overall efficiency for testing (coefficient of efficiency)</b>	<b>Number of RVs (% of training data set)</b>
Multi-day-ahead of multiple reservoir releases	0.89	3.1%
Multi-day-ahead of multiple canal diversions	0.76	2.2%
Multi-hourly-ahead of multiple canal diversions	0.86	0.9%
Multi-day-ahead streamflow and water loss/gain	0.80	2.4%

Table 5.1 also shows the percentage of relevance vectors (RVs) that were used to build each model from the training data set. This percentage of relevant observations ranges from 0.9% to 3.1%. This means that the model ignores a high percentage of observations to avoid over-fitting. This low percentage illustrates that the Bayesian learning procedure embodied in the MVRVM is capable of producing very sparse models. Therefore, we can see an important advantage of using MVRVM models which are capable of reducing model complexity to avoid over-fitting.

The state of a regulated river basin system involves both the physical characteristics of the system and the influence of human behavior in its management. Such systems are characterized by hydrologic processes that are intrinsically nonlinear and nonstationary. In ML modeling, we are training a model to find its parameters and learn the underlying nonlinear functions of dynamic systems based on measured data (which are not free of noise or measurement uncertainty). Moreover, ML theory faces two issues: robust performance with new data and low complexity in model formulation (Mjolsness and DeCoste, 2001). Therefore, this dissertation seeks models that have low

complexity in the explanation of complex real-world water systems (i.e. regulated river basin systems) while maintaining good accuracy and robustness characteristics with respect to future changes in the nature of the input data. Accuracy and low model complexity of the MVRVM model was demonstrated above. Bootstrap analyses (described in Chapter 2, 3 and 4) was used to explore the hypothesis that Bayesian learning machines have the ability to guaranty a robust model. Narrow confidence bounds in the bootstrap histograms imply low variability of the statistics (used to test the model) with changes in the nature of the input data, which indicates that the model parameter set is robust. The bootstrap histograms of the MVRVM models presented in Chapters 2, 3 and 4 show narrow confidence bounds for the root mean square error for the testing phase, which indicate that the MRVM is robust. Also, these bootstrap histograms are much narrower compared to the histograms of artificial neural networks, which are traditional and widely used approaches to machine learning. Therefore, MVRVM models allow more robust performance than ANN models.

This research contributes to the introduction of a sparse Bayesian regression machine that learns the dynamic behavior of regulated river basin system, produces multiple forecasting outputs, and accounts for uncertainties. The versatility of this model is illustrated by its application on a variety of subsystems which are part of a regulated river basin system: multiple reservoirs (Chapter 2), multiple canals (Chapter 3); and streamflow and water loss/gain in a river reach (Chapter 4).

A potential contribution of this dissertation is to apply a data driven model that is based on machine learning as a practical and effective way of analyzing the vast amount of data already recorded in a real automation/internet system (i.e. Sevier River Basin in

Utah), and therefore to exploit the investment that is being made in data acquisition and computerization.

In summary, this dissertation demonstrated the successful robust performance of the MVRVM modeling approach when applied to real-time forecasting of multi-reservoir, multi-canal, and river system operation. The results of this research can have significant potential value in real-time river basin management, allowing operators to more intensively manage the available water resources and make better-informed and timely decisions.

## **5.2 Recommendations and Future Work**

Real-time applications of data driven models that are based on machine learning obviously depend on the accuracy and reliability of sensor data. Inevitable failures on the sensor reading can be related either to a sensor failure or a system failure. Therefore, future research should be directed toward addressing sensor data validation and sensor failure detection for complex river basin systems.

The relevance vectors (RVs) are the summary of the most essential observations of the training data set to build the MVRVM. Future research will be performed by analyzing with more details the statistical and physical meaning of RVs with respect to the training time series data set. For example, this analysis can also be related to whether we would recommend reducing the number of historical data observations for retraining the model with new data in the future.

Application of a hybrid model (e.g. Bayesian approach embedded in ANN model) is being applying in water resources modeling with promising preliminary results. This

hybrid model is being tested and compared to the MVRVM model in terms of accuracy, complexity and robustness.

Future works will be carried out to apply machine learning approaches for optimizing the performance of a physically based model that is currently used on river basin systems.

## **References**

Mjolsness E, DeCoste D (2001) Machine learning for science: state of the art and future prospects. *Science* 293:2051–2055.

Thayananthan A, Navaratnam R, Stenger B, Torr PHS, Cipolla R (2008) Pose estimation and tracking using multivariate regression. *Pattern Recognit Lett* 29(9):1302-1310.

## APPENDICES

## Appendix A. Kernel Functions

### A.1 Gaussian Kernel:

$$K(x, x^{(n)}) = \exp(-r^{-2} \|x - x^{(n)}\|^2)$$

### A.2 Laplace Kernel:

$$K(x, x^{(n)}) = \exp(-(r^{-2} \|x - x^{(n)}\|^2))^{1/2}$$

### A.3 Cauchy Kernel:

$$K(x, x^{(n)}) = 1 / (1 + r^{-2} \|x - x^{(n)}\|^2)$$

### A.4 Cubic Kernel:

$$K(x, x^{(n)}) = (r^{-2} \|x - x^{(n)}\|^2)^{3/2}$$

where  $r > 0$ , is the kernel width parameter, which is a smoothing parameter to define a basis function to capture patterns in the data. This parameter cannot be estimated with the Bayesian approach. For this research, a sensitivity analysis is performed to estimate the kernel width that gives accurate test results.



## Appendix B. Statistics for judging model performance

### B.1 Nash coefficient of efficiency (E):

$$E = 1 - \frac{\sum_{n=1}^N (t_n - t_n^*)^2}{\sum_{n=1}^N (t_n - t_{av})^2}$$

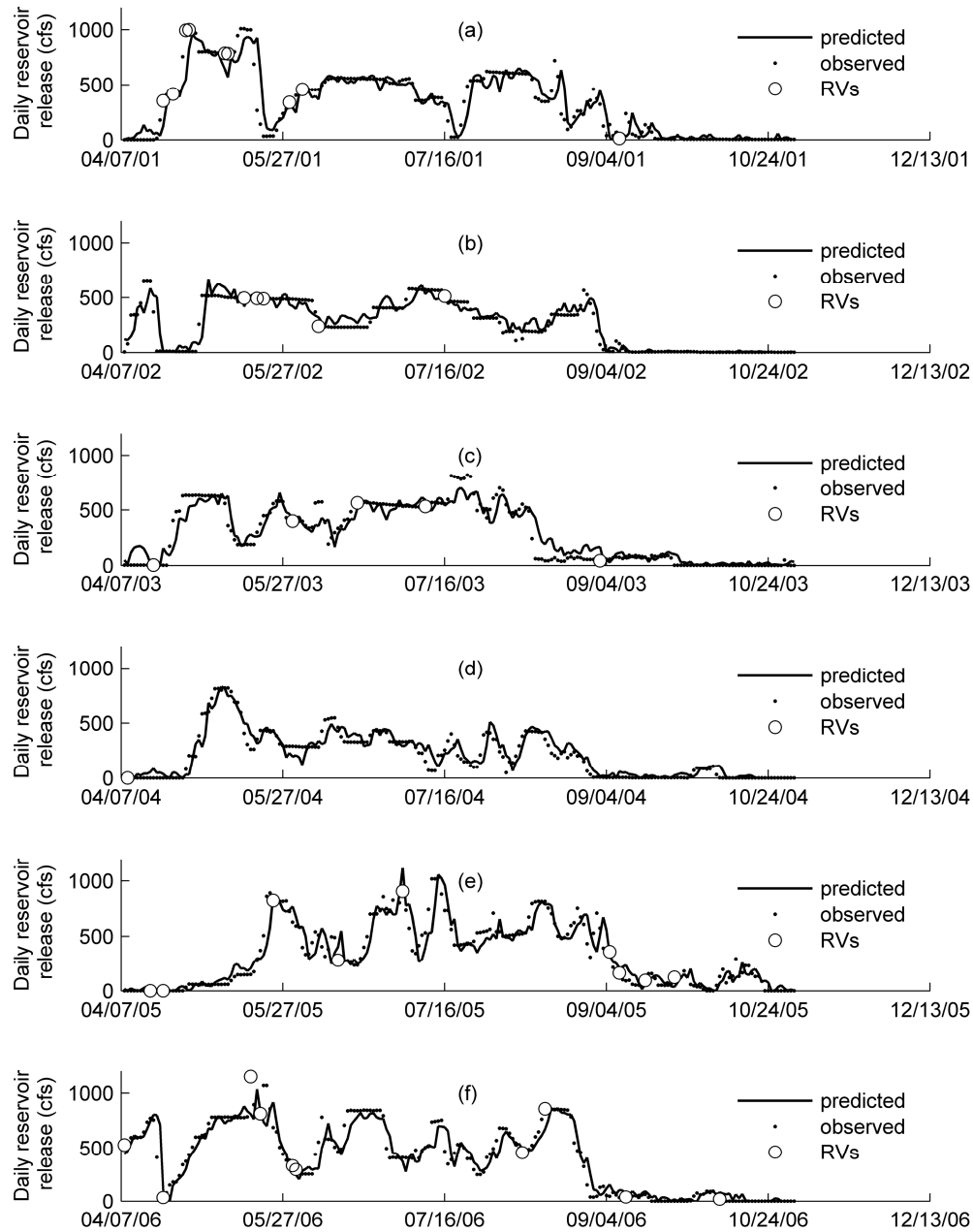
### B.2 Root mean square error (RMSE):

$$RMSE = \sqrt{\frac{\sum_{n=1}^N (t_n - t_n^*)^2}{N}}$$

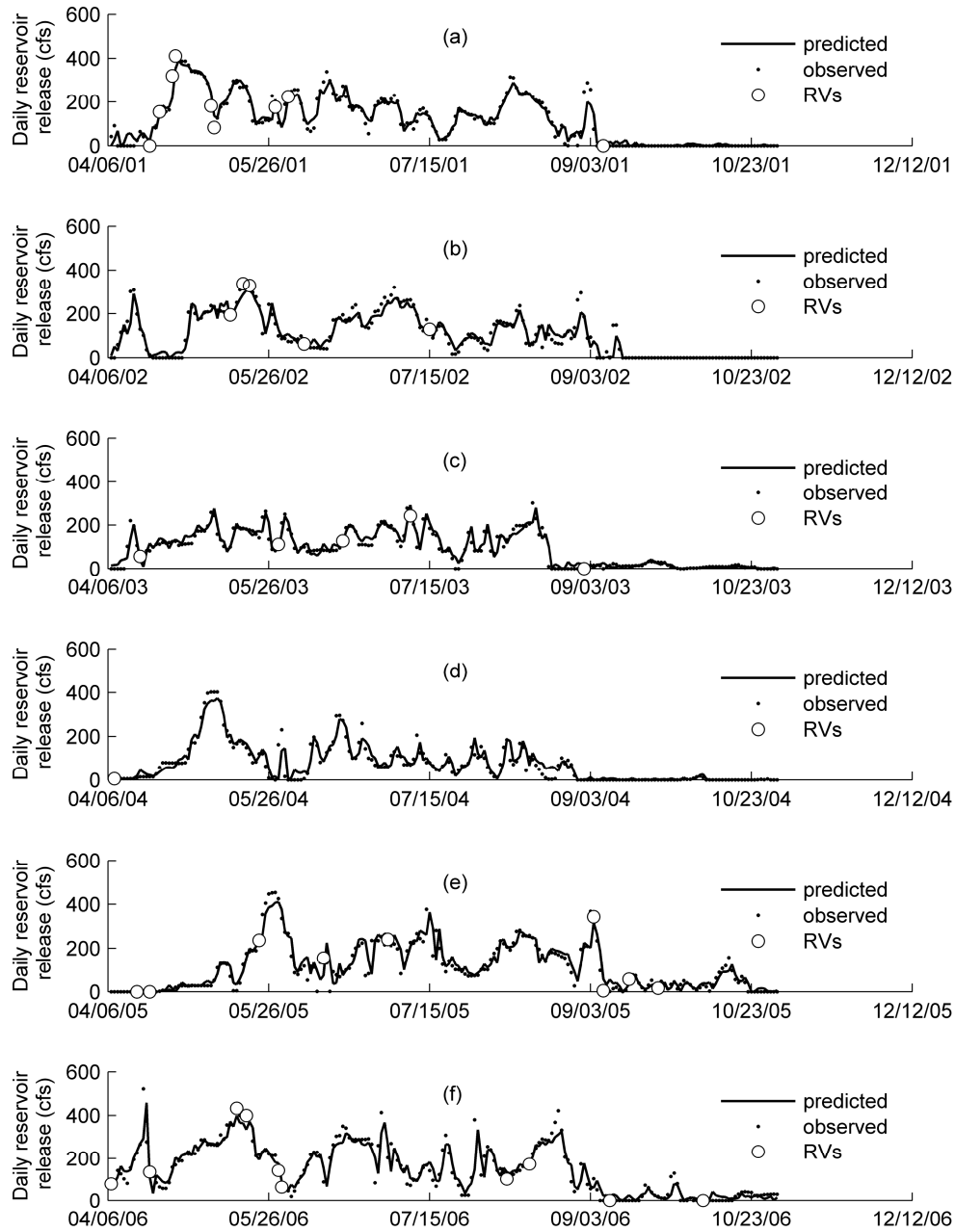
where  $t$  is the observed output;  $t^*$  is the predicted output ;  $t_{av}$  is the observed average output and  $N$  is the number of observations.

### Appendix C. Training phase plots of reservoir release predictions

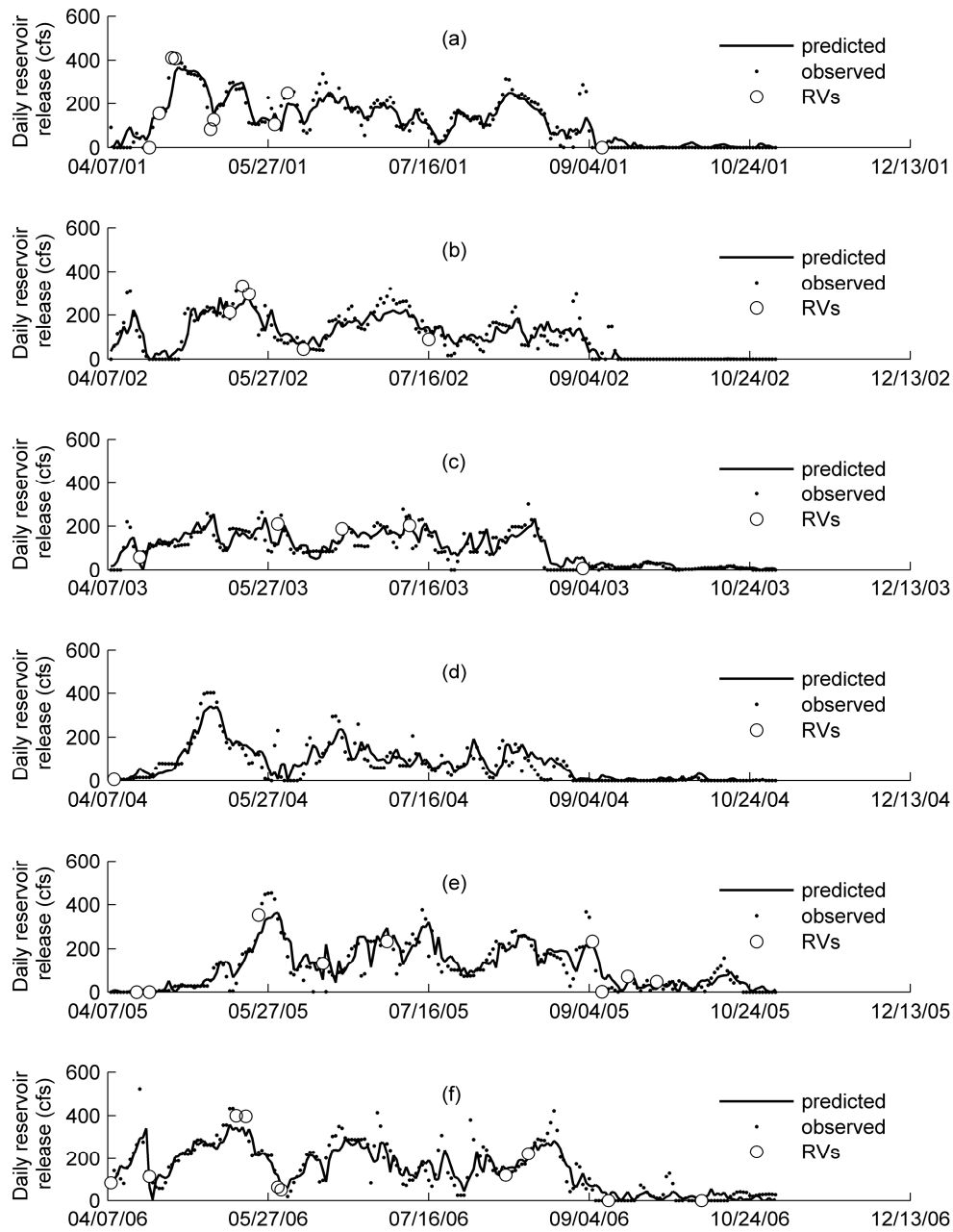
The appendix C to Chapter 2 gives the training phase plots of release predictions for Sevier Bridge reservoir one day ahead (Fig. C.1), and DMAD reservoir one and two days ahead (Fig. C.2 and Fig. C.3, respectively).



**Fig. C.1** Plot of observed versus predicted releases of Sevier Bridge reservoir two days ahead, and RVs of the MVRVM. Training phase (2001(a) – 2006(f) irrigation seasons)



**Fig. C.2** Plot of observed versus predicted releases of DMAD reservoir one day ahead, and RVs of the MVRVM. Training phase (2001(a) – 2006(f) irrigation seasons)



**Fig. C.3** Plot of observed versus predicted releases of DMAD reservoir two days ahead, and RVs of the MVRVM. Training phase (2001(a) – 2006(f) irrigation seasons)

## CURRICULUM VITAE

Andres M Ticlavilca

**EDUCATION**

Ph.D. in Civil and Environmental Engineering (Water Resources), Utah State University. (May 2010). Thesis: “Multivariate Bayesian Machine Learning Regression for Operation and Management of Multiple Reservoir, Irrigation Canal and River Systems” Advisor: Dr. Mac McKee.

M.S. in Irrigation Engineering, (2006), Utah State University. Thesis: "Calibration, Siting, and Design of Flow Measurement Structures in Logan River Irrigation Canals" Advisor: Dr. Gary P. Merkley.

Minor in Agricultural Business Management, (2002), La Molina National Agrarian University, Peru.

B.S. in Agricultural Engineering, (2000), La Molina National Agrarian University, Peru.

**PROFESIONAL EXPERIENCE**

**Graduate Research Assistant**, Utah Water Research Laboratory, May 2006-Present

- Developing data-driven and Bayesian machine-learning models for water resources, hydrologic and hydraulic systems in the Sevier River Basin, Utah.
- Working on the installation of telemetry systems for use with soil moisture probes for real-time collection across a large irrigated area in Delta, Utah.
- Developing data-driven and Bayesian machine learning models for agricultural commodity price forecasting.

**Graduate Student Assistant**, Utah Water Research Laboratory and Biological & Irrigation Engineering Department, January 2006 – April 2006

- Worked in a project for flood and drainage plain mapping in Box Elder County using ARCGIS.

- Worked in a project for the installation of a new water level recording system for flow measurement structures to enhance water delivery service and improve the management of water resources in Cache Valley, Utah.

**Research Graduate Student**, Biological & Irrigation Engineering Department, April 2005 - December 2005

- Worked on water management improvement project in Logan River Irrigation Canals.

**Graduate Student Assistant**, International Irrigation Center, USU, February–March 2005

- Worked on technology transfer on irrigated agriculture and water management for irrigation purposes for participants from different countries.
- Experience in courses and conference arrangements (travel, accommodations, lodgings and organization) in the United States.

**Consultant Engineer**, Ministry of Agriculture, Peru, 2003

- Participated in the formulation of rural development projects which offered technical training to farmers. Projects were focused on economical and financial assessment in production agriculture.

## **LANGUAGES**

Spanish: (Native Language)

English: Speaking, Reading, Writing (Excellent)

## **TEACHING EXPERIENCE**

### **Teacher Assistant**

- Open Channel Hydraulics, CEE 6500. (Fall 2009). Instructor: Dr. Mac McKee.
- Field Techniques, Hydraulic Modeling, and Habitat Assessment. (Jan 10-11, 2009). Master Course at the International Conference of Science and Information Technologies for Sustainable Management of Aquatic Ecosystems. Concepcion,

Chile. Instructor: Dr. Thomas B. Hardy.

### **Guest Lecturer**

- Data-Driven Modeling and Machine Learning Regression Approach in Water Resources Systems (Spring 2009). Water Resources System Analysis, CEE 6410. Instructor: Dr. Bruce Bishop.
- Machine Learning Applications for Integrated Real-Time Water Management (Fall 2009). Division of Environmental Engineering Seminar, CEE 6800. Instructor: Dr. David K. Stevens.

### **Seminar Instructor**

- Seminar on Application of Machine Learning for Water Resources Management (Fall 2009). Utah State University.

### **PRESENTATIONS AND POSTERS AT PROFESSIONAL MEETINGS**

Ticlavilca A, McKee M., Walker W. (Dec 2009). Real-Time Forecasting of Streamflow and Water Loss/Gain in a River System by using a Robust Multivariate Bayesian Regression Model. Poster presented at the 2009 AGU Fall Meeting. Hydrology Section. San Francisco, California.

Ticlavilca A, Feuz D., McKee M. (Jul 2009). Sparse Bayesian Regression Approach to Forecast Agricultural Commodity Prices. Poster presented at the Agricultural & Applied Economics Association 2009 Joint Annual Meeting, Milwaukee, Wisconsin.

Ticlavilca A, McKee M., Walker W. (Jul 2009). Forecasting Irrigation Canal Demand Using Multivariate Sparse Bayesian Learning Approach. Presentation at the AWRA 2009 Summer Specialty Conference. Adaptive Management of Water Resources II. Snowbird, Utah.

Ticlavilca A, McKee M. (Nov 2008). Bayesian Learning Machine Approach to Optimize Operation of a Multiple Reservoir System. Presentation at the AWRA 2008 Annual Conference. American Water Resources Association. New Orleans, Louisiana.

Ticlavilca A, McKee M. (May 2008). A Multivariate Bayesian Machine to Forecast Flow Releases from a Multiple Reservoir System. Presentation at the 2008 Joint Assembly. American Geophysical Union. Fort Lauderdale, Florida.

Ticlavilca A, Tammali B. and Merkley G. (Nov 2005). Water Management Improvement in Cache Valley Irrigation Canals. Paper presented at the 26<sup>th</sup> Annual International Irrigation Technical Conference, 389-404, Phoenix, Arizona.

### **COMPUTER SKILLS**

**Water Resources Engineering Applications:** HEC-1/HEC HMS, HEC-RAS, DAMBRK, HEC-ResSim , SWMM, Visual MODFLOW.

**Spatial Analysis Applications:** ARCGIS, ERDAS.

**Programming Languages:** R, Matlab.

**Design Software:** AutoCAD.

### **AFFILIATIONS**

- American Society of Civil Engineers, American Geophysical Union, American Water Resources Association, Agricultural & Applied Economics Association.